

特 集 「コモンセンス」

言語データからの知識獲得と 言語処理への応用

Knowledge Acquisition from Natural Language Texts and Its Application to Natural Language Processing

井之上 直也

東北大学大学院情報科学研究科

Naoya Inoue

Graduate School of Information Sciences, Tohoku University.

naoya-i@ecei.tohoku.ac.jp, <http://www.cl.ecei.tohoku.ac.jp/~naoya-i>

Keywords: knowledge acquisition, natural language processing, commonsense reasoning.

1. はじめに

本稿では、言語データから獲得できるコモンセンスと、その自然言語処理への応用について解説する。言語処理とコモンセンスの関わりを見るために、自然言語処理の黎明期である 1970 年代から今に至るまでの歴史を、まずは簡単に振り返ってみようと思う。1970 年代は、計算機上に人間のもつ知識をどのように表現するか、また、それらを用いて人間のような常識推論をいかに実現するか、という点でさまざまな研究が行われた。例えば、後に構文解析の分野で多くの業績を上げることになる Eugene Charniak 氏（当時 Brown University）は、物語における登場人物の意図理解の問題に取り組んでいた [Charniak 83]。次の文章が与えられたとする。

(1) *John got a rope. He was going to immobilize Bill.*

John がロープ (*rope*) を用いて次に取り得る行動は何だろうか。この情報は物語においては明示されていないが、我々人間には、*John* は *Bill* をロープで縛ろうとするだろう、と容易に推測できる。ほかに、可能な行動はいくつか考えられるが（例えば、「ロープで荷物を縛る」など）、*John* は *Bill* を拘束しよう (*immobilize*) としていることから、次のもっともらしい行動を絞ることができる。Eugene Charniak 氏は、このような推論を計算機で実現するための知識表現の方法や、その推論機構についての研究を重ねていた。手続き的知識の表現方法であるスクリプト理論 [Schank 75]、自然言語理解を、入力文に対する最良の説明の生成問題（仮説推論）として定式化する Interpretation as Abduction [Hobbs 93] などのもその一例である。

1990 年代後半には、言語処理のための資源として、

WordNet [Fellbaum 98]、FrameNet [Baker 98]、Cyc^{*1} など、さまざまな世界知識のデータベースが、小・中規模ではあるが人手により整備された。しかしながら、冒頭のような言語理解の実現に必要な知識は、膨大である。当時は知識を大規模に用意する方法論がなく、実問題にスケールさせるのは困難であった。

2000 年代前半、大きな変化が訪れる。それまでに蓄積されてきた電子新聞記事、Weblog などの言語データが大規模に利用可能になったのである（いわゆるビッグデータ時代の幕開けである）。これらの言語データには、「業績が悪化する→株価が落ちる」などの物事の因果関係や、「徹夜する→肌荒れする」などの人々の経験が自然言語で記述されている。このような性質に着目し、これらの言語データから、自然言語処理のための知識を大規模に獲得する研究が一気に加速した。例えば、同義語・上位下位語などの名詞間の意味関係 [Pantel 06, Snow 05]、X commit a crime → X is convicted などの事象間の関係知識 [Chambers 08, Chklovski 04a, Schoenmackers 11, Shibata 11]、is capital of (Tokyo, Japan) などの固有名詞間の関係知識 [Fader 11, Yates 07] など、さまざまな知識の自動獲得について研究がなされた。これらのアプローチは基本的には、“X such as Y” といった語彙統語パターンを用いて同義関係となる単語ペア、または共参照項を手掛かりとして関連のある事象のペアを収集し、Pointwise Mutual Information などの統計的な尺度を用いてペアの関連の強さを見積もるといったものである。

また、2010 年代に入ると、集合知を利用した知識獲得の動きも活発化した。Web 上で発信する情報を計算機にとって扱いやすいものとする枠組みである Linked Open Data (LOD) を基盤として、例えば、クラウドソーシング・オンラインユーザ協調により、因果関係などのさまざまな種類の知識を収集した ConceptNet [Speer 12]、

*1 <http://www.cyc.com/>

Wikipediaの構造化情報(LODの一種)をもとに構築した固有名詞に関する知識ベースであるDBpedia^{*2}, FreeBase, wikidata^{*3}などがあげられる。

このような大規模知識獲得が可能になったことにより, 再び冒頭のような推論を目指す研究が行われるようになり, 今日に至る。このような背景のもと, 本稿では, 自然言語処理におけるコモンセンス(知識)に関する最新の研究動向の中から, 著者が重要と考えるいくつかのホットトピックを紹介する。具体的には, 連続空間への知識の埋込み, 埋込み空間上での論理的推論, および常識推論の性能評価を行う最近のベンチマークデータに焦点を当ててその研究動向を紹介する。紙数の都合上, これらすべてを詳細に説明することはできないため, 研究全体の流れを把握できるような解説となるよう努めた。個々の技術の詳細についてさらに興味のある読者は, ぜひ引用文献を辿ってほしい。

2. 連続空間への知識の埋込み

言語データから獲得された知識には, 同じ概念を表現するさまざまな同義表現や, 個別の経験から一般的な法則まで, さまざまな粒度の知識が書かれている。したがって, 獲得した知識を有効に言語処理に活用するためには, 獲得した知識を適切にまとめ上げ, 汎化しておかなければならない。そこで2010年代半ばより, 獲得した知識を連続空間上に埋め込むことで汎化を達成しようとする, 知識埋込み(Knowledge Embedding)の研究^{*4}が盛んに行われ, 一大トレンドとなっている。本章では, これらの最新の研究動向に触れていく。

2.1 固有名詞間の関係知識

知識の埋込みの研究の中で最も盛んに行われているのは, 固有名詞の関係知識である。最も基本的なモデルの一つは, TransEと呼ばれるものであり[Bordes 13], ここでそのアイデアを説明する。

まず, 固有名詞間の関係知識をエンティティ h, t とそれらの関係 r からなる三つ組 (h, r, t) により表現する。例えば, 「東京」が「日本」の「首都である」という関係を $(\text{Tokyo, is_capital_of, Japan})$ と表す。また, それぞれのエンティティと関係を n 次元実数空間上の点(埋込み)として表現する。TransEの基本的なアイデアは, 知識ベースに書かれた関係 (h, r, t) について, 対応する埋込み表現 h, r, t を, 下記の損失関数を最小化するように学習することである:

$$\sum_{(h, r, t) \in K^+} \sum_{(h', r', t') \in K^-} [\gamma + f(h, r, t) - f(h', r', t')]_+ \quad (1)$$

ここで, γ はマージン, K^+, K^- はそれぞれ知識ベースに含まれる三つ組の集合(正例), 知識ベースに含まれない(偽の)三つ組の集合(負例)であり, $[\cdot]_+ = \max(0, \cdot)$ である。また, $f(h, r, t) = -\|h + r - t\|_2$ (またはL1ノルム)であり, 三つ組 (h, r, t) の妥当性をその埋込み表現に基いて測る関数である。学習が進むと, 知識ベースに存在する三つ組 (h, r, t) については, $h + r$ が t に近づくようになる。

こうして学習された分散表現は, 例えば“Where was Obama born?”といった質問応答に利用することができる。具体的には, エンティティ *Obama*, 関係 *is_born_in* に対応するベクトル表現をそれぞれ **obama, is_born_in** とし, 知識ベース上のあるエンティティ t の埋込み表現を t とする。質問応答を行うには, $f(\text{obama, is_born_in, } t)$ を最大化するようなエンティティ t を探せばよい。表層的なパターンマッチングを行う場合に比べ, 訓練データに $(\text{Obama, is_born_in, } \cdot)$ という関係が存在していなくても, 似た関係, 例えば $(\text{Obama, is_given_birth_in, } \cdot)$ などが存在していれば回答を導き出せるなど, 柔軟な推論が可能になるという利点がある。

その後, 関係の表現方法, 多対多の関係性の表現など, さまざまな観点からいくつかの発展系が現在進行形で編み出されている。そのすべてを網羅することは本稿の趣旨から外れるので, さらに興味のある読者には, 知識埋込みに関するサーベイ論文[Cai 18, Wang 17]を読むことをお勧めする。

2.2 事象間の因果関係知識

固有名詞間の関係のほか知識の埋込みの研究対象とされているのが, 事象間の因果関係知識, 特にスクリプト的知識(Script Knowledge)である。スクリプトとは, 同時に起こり得る典型的な事象の順序付き集合である[Schank 75]。例えば, 「レストラン」のスクリプトには, 「椅子に座る」, 「メニューを見る」, 「注文する」といった事象の集合が含まれる。本稿の冒頭の例からもわかるように, このような知識は高度な言語処理を行ううえで大変重要である。

1章でも触れたように, 2000年代, こうした知識をWeb上の大規模な文書集合から獲得する手法が盛んに研究された。典型的な事象間関係の自動獲得の手法は, “, and then” や “because” などの語彙統語パターンを用いるもの[Abe 08, Chklovski 04, Luo 16], 共参照関係を手掛かりとして用いるもの[Chambers 08, Chambers 09]が主流であった。これらを手掛かりを用いて, コーパス上の, 例えば “John touched Mary and then kissed her.” という記述から事象間の関係知識を大規模に獲得し, Pointwise Mutual Informationなどの統計的尺度を

*2 <http://wiki.dbpedia.org/>

*3 <https://www.wikidata.org/>

*4 自然言語処理における埋込み一般についての研究は, [岡崎16]を参照されたい。

用いて因果の強さを推定していた。

こうした自動獲得のアプローチにおける大きな課題の一つとして、コーパスから大量に獲得した因果関係の事例をどの程度汎化し知識とするか、という問題がある。例えば、“*John was fined because he smoked in a non-smoking hotel room.*” “*Mary smoked in a non-smoking room, so she was fined 10,000 yen.*” という因果関係の記述があったとしよう。これらの因果関係の事例を汎化した知識として、例えば「人が禁煙室 (*non-smoking room*) で喫煙する→罰金を払わされる」という因果関係が考えられる。一方で、理屈上はこれをさらに汎化し、「人が喫煙する→罰金を払わされる」という因果関係を考えることもできる。理想的には前者が良さそうに見えるが、これを定量的に測るにはどうしたらよいらうか。先行研究では、汎化の粒度を「動詞のみ」、「主語、動詞、目的語」などに一律固定する、というアプローチが取られた。しかし、これでは当然、後者のような妥当性の低い知識も獲得されてしまう。一方で、汎化をしないと特殊すぎる知識が得られてしまう。

この問題を解消するために、近年は連続空間にスクリプト知識を埋め込み、連続空間上で因果関係の対応を学習する研究が盛んに行われている [Granroth-wilding 16, Liu 16, Modi 14, Pichotta 16a, Weber 17, etc.]. これらの手法では、2.1 節で解説した固有名詞間の関係知識の埋込みと同様に、 n 次元実数空間上で事象を表現し、因果関係の推定が正しく行えるような事象の埋込み表現を自動的に学習する。これはすなわち、記号表現から埋込み表現へのマッピングを通して、どの情報を捨象するかを決めている一すなわち事象の汎化とみなすこともできる。例えば、*fined 10,000 yen* と *fined yesterday* という表現を含む因果関係において、*10,000 yen* と *yesterday* が因果関係の表現に重要でないならば、これら二つの事象は埋込み表現上では同一のものになることが期待される。

埋込みの基本的なアプローチは、先に述べたような手法により因果関係の事例を大規模に獲得し、獲得した因果関係を正しく識別できるような埋込み表現を学習することである。これまでに、事象を構成する単語の情報をどのように組み合わせて事象の埋込み表現を生成するか、因果関係認識のアーキテクチャをどのように構成するか、といった方向性からさまざまな検討が行われている。事象の埋込み表現という観点では、重み付き加法 [Granroth-wilding 16], リカレントニューラルネットワーク [Pichotta 16a], テンソルに基づく述語表現 [Weber 17] などが検証されている。また、アーキテクチャの観点では、入力を二つの事象、出力をそれらの事象間の因果関係の有無を表すスコアとする Siamese Networks [Granroth-wilding 16, Liu 16, Modi 14, Weber 17], Sequence-to-Sequence モデル [Sutskever 14] に基づくモデル [Pichotta 16a, Pichotta

16b, Roemmele 17], TransE の変形 [Zhao 17] が存在する。また、word2vec [Mikolov 13] を因果関係記述の集合に適用し、因果関係を予測するのに特化した分散表現 (Causal Embedding) を学習する手法 [Sharp 16] もある。また、一方で、知識を連続空間に埋め込まず、知識の汎化の良さを統計的尺度に基づいて決める手法 [Yokoi 17] も提案されている。

3. 埋込み空間上での論理的な推論

得られた知識を言語処理の中で有効活用するためには、これまでに述べた汎化の課題の解決に加えて、さらに知識をうまく使いこなすための推論機構が必要になる。本稿では、埋込み空間上での論理的推論を実現する、筆者が注目している最近の三つの取組みに触れる。

3.1 Natural Logic

Natural Logic [Maccartney 09] は、二文 T, H が与えられたとき、自然言語表現のうえで推論を行いながら、 T が H を含意するかを判定するための枠組みである。基本的には、まず T を H に書き換える手順を、「単語の挿入」、「単語の削除」、「単語の置換」という 3 種類の編集演算子の系列により表現する。ある演算子のある単語に適用した際、編集前の文と編集後の文の意味関係がどのように変化するかがあらかじめ規定されており、最終的にこれらの意味関係を総合的に見て、 T と H がどのような関係にあるかを決定する。

例えば、文 T : “*John has a dog.*” と文 H : “*John does not have an animal.*” の関係を判定してみよう。まず T に対して「*dog* を *animal* に置換する」という操作を行い、文 T' : “*John has an animal.*” を得る。次に、Natural Logic のルールより、文間の意味関係が決まる。 $\text{dog} \sqsubset \text{animal}$ なる関係をもつ単語対を用いた置換操作なので、 T は T' を含意する関係にある、と結論付けられる。同様に、編集・意味関係の推論を行う。 T' に対して「*have* を否定」し、 T'' : “*John does not have an animal.*” (= H) を得る。not の挿入により、 T' と T'' の含意関係は成立しなくなり、排他的な (negation) 関係となる。また、もとの T と T'' の関係は、代替 (alternation) の関係となる。ゆえに、 H は T により含意されない、と判定できる。

[Bowman 15b] は、このような Natural Logic に基づく含意関係認識を連続空間上で実現する手法を示した。より具体的には、句間の意味関係をより頑健に同定するために、句を再帰ニューラルネットワークにより連続空間に埋め込み、これを入力として句間の関係を同定する分類器を構築した。この手法により、含意関係認識のベンチマークセット SICK [Marelli 14] の上で、76.9% の精度で含意関係認識ができることを示した。Natural Logic では、否定や量子子といった演算子を取り扱うこ

ともでき、柔軟な論理推論の実現の大きな一歩を踏み出した成果であるといえる。

3.2 Dependency-based Compositional Semantics

Dependency-based Compositional Semantics (DCS) [Liang 13] は、質問応答への応用を意識してつくられた自然言語文の意味表現の一種である。DCSでは、文の意味はデータベースクエリ的に定義される。より具体的には、その自然言語のデータベースクエリ表現にあたる DCS 木（おおむね単語間の依存構造にあたる）と、その問合せ結果の集合（外延）により、自然言語の意味を規定する。

例えば、“banned drugs”の意味は、 $\text{ban} \xrightarrow{\text{COMP}} \text{drug}$ という DCS 木と、その問合せ結果 {marijuana, heroin, ...} で規定する。問合せ結果は、各構成要素の問合せ結果を、依存構造に基づいて順に計算し、積集合をとったもの、と考えることができる。例えば、前述の問合せ結果は、drug の問合せ結果の集合（例えば、{sleeping pill, marijuana, ...}）と、ban されるものの問合せ結果の集合（例えば、{murder, marijuana, ...}）の積集合を取ったもの、と考えることができる。

[Tian 14] は、DCS を用いた論理推論の枠組みを提案し、さらに [Tian 16] は、クエリへの問合せ結果を連続空間に埋め込むことにより、DCS をより柔軟に表現する手法を示した。具体的には、まず自然言語文の各単語に埋込み表現 v (ベクトル)、DCS 木の依存関係に行列 M を割り当てる。また、DCS 木での各ノードの問合せ結果を、依存関係行列によるベクトルの変換により連続空間上に埋め込むことで表現する。例えば、先ほどの例を具体的に計算してみよう。まず、「ban されるもの」の問合せ結果は v_{ban} 、drug の問合せ結果は $v_{\text{drug}} M_{\text{COMP}}$ と表現する。最後に、これらの問合せ結果の積集合を計算するために、ベクトル間の足し算を行う。すなわち、“banned drug”の問合せ結果は、 $v_{\text{ban}} + v_{\text{drug}} M_{\text{COMP}}$ の近傍にある単語、と表現できる。なお、これらの埋込みは、依存構造解析済みのコーパスを用いて学習する。この結果を [Tian 14] と組み合わせることで、DCS を用いた柔軟な論理推論の実現が可能になり、今後の展開が楽しみである。

3.3 SLD 導出に基づく定理証明

[Rocktäschel 17] は、ホーン節論理を知識表現に用いた、Selective Linear Definite clause (SLD) 導出に基づく定理証明を埋込み空間上で実現する方法を示した。SLD 導出では、与えられたゴールを、知識ベースのヘッドとのパターンマッチングにより次々と証明していくが、この際に意味的に類似した述語のマッチングを考慮できない。例えば、grandfatherOf (John, Bob) と、grandpaOf (John, Bob) は、述語の意味が似ていても、表層的な違いによりマッチングを取ることができない。

そこで [Rocktäschel 17] は、スコア付けられたパターンマッチングを埋込み空間の上で柔軟に行い、最終的にはゴールを証明できるか否かでなく、(パターンマッチングのスコアに比例する) 証明成功のスコアを返すような定理証明器を提案した。

具体的には、まず知識ベースに現れる述語と定数にベクトル表現を割り当て、述語と定数を連続空間に埋め込む。ゴールと知識ベースのパターンマッチングの際には、これらのベクトル表現を用いて類似度を計算し、証明を続けていく。このままでは証明は無限に続いてしまうので、ある一定の深さ d で証明を打ち切る。述語と定数のベクトル表現は、知識ベースから証明できる（できない）ゴールの証明スコアを高く（低く）するようにベクトル表現を学習する。評価では、この定理証明器を、2.1 節で説明したような知識ベース補完のタスクで評価している。別のリンク予測器と本提案を組み合わせることにより、柔軟な論理的な推論を行いながらのリンク予測が可能となり、さらに予測性能を向上させられることを示している。このように、一般的な定理証明を連続空間上に埋め込むことで、論理表現を用いた高度な知識表現のうえでさまざまな推論が可能になる。

4. 常識推論のベンチマーク

近年、計算機が知識に基づく言語処理を適切にできているかを試すためのさまざまなベンチマークが提案されている。本章では、このうちのいくつかを例にとって紹介する。

4.1 含意関係認識

常識推論のベンチマークとして古典的なのは、二文 T , H が与えられたときに、 T が H を含意するか (T が真のとき、 H も真といえるか) を認識する問題である。例えば、下記の 2 文が与えられたとする。

● T : *Cavern Club sessions paid the Beatles £15 evenings and £5 lunchtime.*

● H : *The Beatles perform at Cavern Club at lunchtime.*

Cavern Club sessions が *the Beatles* に対してランチタイム (*lunchtime*) に £5 を支払った (*paid*) ということから、*The Beatles* が *Cavern Club* で演奏した (*perform*) ということが推論でき、「含意する」が正解である。含意関係認識の問題は、Bar-Ilan 大学の Ido Dagan らの研究グループの主導により、2006 年に Pascal RTE Challenge として共通タスク化され [Dagan 06]、過去 7 回の評価型ワークショップが行われた。

最近では、スタンフォード大学の研究グループより、クラウドソーシングに基づいて大規模に構築された含意関係認識のデータセット Stanford Natural Language Inference (SNLI; 57 万事例、画像の説明文ドメイン)

[Bowman 15a], Multi-NLI (約 43 万事例, さまざまなドメイン) [Williams 18] がリリースされた. これは過去にリリースされた RTE のデータに比べて相当に大規模で, 多くの研究者が含意関係認識モデルの研究に取り組むきっかけをつくった [Gong 18, Parikh 16, Zhao 16, etc.]. RTE Challenge の後継として, 2 文の「含意」関係でなく「類似」関係の判定をタスクとする, **Semantic Textual Similarity (STS)** というタスクも提案され, 2012 年より **SemEval** のタスクの一つとして採用されている [Agirre 16].

冒頭から述べているように, この種の問題を解くために必要な知識は無数にあると考えられるため, 少数の訓練データからそれらを学習できることはほとんど期待できない. 特に, 含意関係認識の初期の頃は訓練データが数百事例のオーダであったため, いかに外部の知識を取り込み, 利用するかという点でさまざまな研究がなされていた [Bos 06, Ovchinnikova 11, Raina 05]. しかし, 大規模な SNLI がリリースされると, 研究の観点は, 外部知識を使わずに, いかに深層学習モデルのアーキテクチャを洗練するかという点に集中している.

4.2 ストーリー予測

南カリフォルニア大学の研究グループは, 因果関係モデルを評価するためのベンチマークとして, **Choice of Plausible Alternatives (COPA)***5 という常識推論問題を提案した [Roemmele 11]. COPA は, 前提 P と二つの文 A_1, A_2 (alternatives と呼ばれる) が与えられたとき, P の結果 (または原因) としてふさわしい文を選ぶ問題である. 例えば, 下記の問題を見てみよう.

● P : *The man broke his toe. What was the CAUSE of this?*

● A_1 : *He got a hole in his sock.*

● A_2 : *He dropped a hammer on his foot.*

つま先 (*toe*) を怪我した (*broke*) ことの原因としては, 靴下 (*sock*) に穴が空いたから (*got a hole*) ではなく, ハンマー (*hammer*) を足 (*foot*) の上に落としたから (*dropped*), ということがよりふさわしい. つまり, A_2 が正解である. 著者の Web サイトにおいて, データセット 1000 問 (開発データ, テストデータそれぞれ 500 問) が一般公開されており, 多くの研究者がこれに挑戦している.

また, 2016 年には, Rochester 大学の研究グループが, COPA を拡張した **Story Cloze Test***6 という問題を提案した [Mostafazadeh 16]. **Story Cloze Test** は, 4 文からなるストーリーの文脈 C と, そのエンディングの候補 A_1, A_2 が与えられたとき, 最も適切なエンディングを選ぶ問題である. 例えば, 下記の問題を考えてみよう:

● C : *Karen was assigned a roommate her first year of college. Her roommate asked her to go to a nearby city for a concert. Karen agreed happily. The show was absolutely exhilarating.*

● A_1 : *Karen became good friends with her roommate.*

● A_2 : *Karen hated her roommate.*

ここでは, A_1 が正解である. 特筆すべき点は, その規模であり, クラウドソーシングを利用して 10 万ストーリーからなるデータセットを一般公開している. また, **Story Cloze Test** を共通タスクとしたコンペティションも開かれ, 2017 年 4 月に自然言語処理のトップ会議の一つである **EACL** のワークショップとして, 各種システムと関連研究の発表が行われた.

現状行われている研究の主な解法は, 因果関係を表すキーワード (“because” など) や照応関係などの手掛かりを用いて, 大規模な文章の集合から常識的な知識を獲得し, これらをもとに 2 文間の因果関係を統計的に計算する手法である [Luo 16]. また, 獲得した知識を **Sequence to Sequence** 学習モデル [Sutskever 14] に投入し, ストーリーの生成器を構築するアプローチもある [Roemmele 17]. 本稿執筆時点 (2018 年 2 月) では, **COPA**, **Story Cloze Test** とともに, まだ 7 割程度の精度でしか解析ができておらず, これからの発展が楽しみな分野である.

4.3 Adversarial Examples

New York 大学の研究グループと Toronto 大学の Hector Levesque の研究グループは, 「統語的手掛かりなどの表層の手掛かりだけでは解けない照応解析の問題を常識推論のベンチマークとして提案した (詳しくは, [Levesque 11] の 3 原則を参照されたい). このテストは, 人工知能研究者 Terry Winograd にちなんで, **Winograd Schema Challenge (WSC)***7 と名付けられた. 下記の例を見てみよう.

(2) *The city councilmen refused the demonstrators a permit because they feared violence.*

(3) *The city councilmen refused the demonstrators a permit because they advocated violence.*

(2) では, *they* の指示対象は *the city councilmen* であるのに対して, (3) では *the demonstrators* であり, これを正しく当ててるのがタスクである.

これらの問題を正しく解くためには, COPA や **Story Cloze Test** と同様, 「ある人が何かを **fear** すると, **refuse** する」といった常識的な因果関係の知識に基づいた予測モデルを構築する必要がある. 当然, 表層的な手掛かりから指示対象を推測するモデルを構築することはできるが, (2) と同時に (3) のような問題が含まれているた

*5 <http://people.ict.usc.edu/~gordon/copa.html>

*6 <http://cs.rochester.edu/nlp/rocstories/>

*7 <https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WS.html>

め、問題文をしっかりと理解できていないと、正解率は低くなってしまふ。常識的な知識を使いこなして初めて高得点が出せる問題集となっている点がポイントである。

2016年には、人工知能のトップ会議であるIJCAIのワークショップとしてWSCの第1回コンペティションが、2018年にはAAAIの一部としてコンペティションが開かれている。2016年のコンペティションにおける優勝者は、2・2節でも触れた、因果関係埋込みに基づくアプローチ[Liu 17]である。しかし、その性能は5割程度であり、まだまだ発展途上の段階である。

こうした敵対的な(adversarial)問題をベンチマークとすることの重要性が近年認識され、一大ムーブメントになりつつある。スタンフォード大学の研究グループ[Jia 17]の事例を紹介しよう。例えば、下記のようなパッセージと質問が与えられたとする。

●Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager.

●What is the name of the quarterback who was 38 in Super Bowl XXXIII?

質問への正しい答えは、John Elwayである。では、このパッセージに次の文を加えたとして：

(4) Jeff Dean is the name of the quarterback who was 37 in Champ Bowl XXXIV.

当然、答えはJohn Elwayのままである。しかしながら、[Jia 17]は、既存のいくつかの質問応答システムの性能をこのような改変前後のパッセージで比べたときに、その性能がF値で20~30ポイント程度、どのシステムも軒並み下がってしまうことを発見した(人間の正答率は誤差程度しか変わらない)。

この発見から、大きく二つのことがいえる。第一に、システムが学習していることは訓練データ特有の傾向であり(回答は最後の人名であることが多いなど)、本当に質問への回答の仕方を学習しているわけではないかもしれない、ということだ。第二に、ベンチマークデータをつくったならば、それが計算機の何の能力を試せるのか、しっかりと検証しておく必要があるということだ。既存のデータセットに対するこうした分析の試みはすでに行われており、興味のある読者は[Sugawara 17]を参照されたい。

敵対的問題を考えることの動機は、計算機が本当に“intelligent behavior”をしているのかを試したい、ということから来ている[Jia 17]。これの一部は、冒頭で述べたような知的な振舞いを実現したいということであり、これはまさに常識推論のベンチマークになってい

る。こうした考え方は業界全体にも広まりつつあり、自然言語処理のトップ会議EMNLP, NAACLにおいても、こうした問題意識をもったワークショップが採択され、開催されている[Ettinger 17]^{*8,*9}。こうした「モデルの知的振舞い、常識推論モデルの良さ」を定量的に測れるような環境が整備されてくると、本稿で紹介したような研究は業界全体でも一気に加速するだろう。

5. おわりに

本稿では、言語データからのコモンセンス獲得、利用というテーマのもとで最新の研究動向を紹介した。その歴史を振り返ってみると、手書きで書いていた知識は自動獲得の方法論により大きく補強され、さらに連続空間への埋込み、および埋込み空間上での柔軟な推論機構の研究が行われるなど、冒頭で述べたゴール—計算機による常識推論の実現—に着実に向かっているように見える。また、計算機が知的な振舞いを行っているかのベンチマークテストについての研究も進み、その成果を定量的に確かめられる環境が徐々にそろってきており、今後こうした研究がより一層増えることが期待される。

「Googleのネコ」に代表される、2010年代に起こった深層学習によるブレイクスルーは、自然言語処理の研究にも大きな影響を与えた。自然言語処理の基礎解析技術を含め、意味解析、文脈解析といった、あらゆる自然言語処理の研究も深層学習化が進んだ。また、従来のように形態素解析・語彙統語解析などを行わず、深層学習の枠組みのもとで、自然言語文を入力として直接タスクを解くアプローチ(end-to-end)の研究も多く登場した。このような状況で、本稿のような方向性の研究がどのように生きてくるのだろうか、という疑問をもつ読者も少なくないだろう。この問いに対する著者なりの考えを最後に述べて、本稿を締めくくろうと思う。

まず、言語処理に対する深層学習の効果は画像や音声分野ほどのインパクトがなく、性能の向上幅は限定的なのが現状である。著者は、その原因を次のように考察する。インパクトの大きかった画像や音声と言語は、入力データの質が異なっており、問題の難しさが異なる。画像や音声は、それだけで自己完結した“アナログ的な入力”である。つまり、そもそもそこからどのような特徴量を取り出すか、ということが大問題である。これに対して、言語は人々の知識に依拠した記号の列であって、それだけで完結するものではない。仮に、人間が表した意味の電気信号(脳波かもしれない)があるとする、言語はある種、その特徴量と考えることもできる。深層学習のキモが特徴量の自動学習であることを考えると、言語処理に真のブレイクスルーが起こるには、あともう

*8 <https://bibinlp.umiacs.umd.edu/>

*9 <https://newgeneralization.github.io/>

一步, 別の何かが必要である。

著者としては, 本稿で見てきたような, 人がもつ知識をいかに獲得し汎化するか, それらを入力テキストから呼び起こし活用する計算機構をいかにつくるか, といった部分がやはり本質的な課題であり, これを追求することが一つのブレークスルーにつながると考えている。さらに, このような知識に基づく堅実なアプローチを粘り強く追求していくことで, 4.3節で述べたような敵対的問題にもだまされにくいモデルが, 自然と出来上がっていくのではないかと考えている。

◇ 参 考 文 献 ◇

- [Abe 08] Abe, S., Inui, K. and Matsumoto, Y.: Two-phased event relation acquisition: Coupling the relation-oriented and argument-oriented approaches, *Proc. 22nd Int. Conf. on Computational Linguistics (Coling 2008)*, No. Aug., pp. 1-8 (2008)
- [Agirre 16] Agirre, E., Banea, C., Cer, D., Diab, M., Gonzalez-Agirre, A., Mihailescu, R., Rigau, G. and Wiebe, J.: Semeval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation, *Proc. SemEval.*, pp. 497-511 (2016)
- [Baker 98] Baker, C. F., Fillmore, C. J. and Lowe, J. B.: The Berkeley Framenet Project, *Proc. COLING*, pp. 86-90, Association for Computational Linguistics (1998)
- [Bordes 13] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J. and Yakhnenko, O.: Translating embeddings for modeling multirelational data, *Advances in Neural Information Processing Systems*, pp. 2787-2795 (2013)
- [Bos 06] Bos, J. and Markert, K.: Recognising textual entailment with robust logical inference, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 3944 LNAI, pp. 404-426 (2006)
- [Bowman 15a] Bowman, S. R., Angeli, G., Potts, C. and Manning, C. D.: A large annotated corpus for learning natural language inference, *Proc. 2015 Conf. on Empirical Methods in Natural Language Processing, Lisbon*, No. September, pp. 632-642, Portugal, Sept. 17-21, 2015 (2015)
- [Bowman 15b] Bowman, S. R., Potts, C. and Manning, C. D.: Recursive neural networks can learn logical semantics, *Proc. 3rd Workshop on Continuous Vector Space Models and Their Compositionality (CVSC)*, pp. 12-21, Beijing, China, July 26-31, 2015 (2015)
- [Cai 18] Cai, H., Zheng, V. W. and Chang, K. C.-C.: A Comprehensive survey of graph embedding: Problems, techniques and applications, *IEEE Trans. on Knowledge and Data Engineering (TKDE'18)* (2018)
- [Chambers 08] Chambers, N. and Jurafsky, D.: Unsupervised learning of narrative event chains, *ACL*, pp. 789-797 (2008)
- [Chambers 09] Chambers, N. and Jurafsky, D.: Unsupervised learning of narrative schemas and their participants, *ACL*, pp. 602-610 (2009)
- [Charniak 83] Charniak, E.: Passing markers: A theory of contextual influence in language comprehension, *Cognitive Science*, Vol. 7, No. 3, pp. 171-190 (1983)
- [Chklovski 04] Chklovski, T. and Pantel, P.: VerbOcean: Mining the web for fine-grained semantic verb relations, Lin, D. and Wu, D., eds., *Proc. EMNLP*, pp. 33-40, ACL (2004)
- [Dagan 06] Dagan, I., Glickman, O. and Magnini, B.: The PASCAL recognising textual entailment challenge, *Machine Learning Challenges, Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment*, pp. 177-190, Springer (2006)
- [Ettinger 17] Ettinger, A., Rao, S., Daumé, H. and Bender, E. M.: Towards linguistically generalizable NLP systems: A Workshop and shared task, *Proc. of the 1st Workshop on Building Linguistically Generalizable NLP Systems*, pp. 1-10 (2017)
- [Fader 11] Fader, A., Soderland, S. and Etzioni, O.: Identifying relations for open information extraction, *Proc. Conf. on Empirical Methods in Natural Language Processing*, pp. 1535-1545, Association for Computational Linguistics (2011)
- [Fellbaum 98] Fellbaum, C.: *WordNet: An Electronic Lexical Database*, MIT Press (1998)
- [Gong 18] Gong, Y., Luo, H. and Zhang, J.: Natural language inference over interaction space, *Int. Conf. on Learning Representations* (2018)
- [Granroth-wilding 16] Granroth-wilding, M. and Clark, S.: What happens next? Event prediction using a compositional neural network model, *AAAI* (2016)
- [Hobbs 93] Hobbs, J. R., Stickel, M. E., Appelt, D. E. and Martin, P.: Interpretation as abduction, *Artificial Intelligence*, Vol. 63, No. 1-2, pp. 69-142 (1993)
- [Jia 17] Jia, R. and Liang, P.: Adversarial examples for evaluating reading comprehension systems, *Proc. 2017 Conf. on Empirical Methods in Natural Language Processing*, pp. 2021-2031 (2017)
- [Levesque 11] Levesque, H. J.: The Winograd schema challenge, *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning* (2011)
- [Liang 13] Liang, P., Jordan, M. I. and Klein, D.: Learning dependency-based compositional semantics, *Computational Linguistics*, Vol. 39, No. 2, pp. 389-446 (2013)
- [Liu 16] Liu, B. and Lane, I.: Attention-based recurrent neural network models for joint intent detection and slot filling, *Proc. Annual Conf. of the Int. Speech Communication Association, INTERSPEECH*, No. 1, pp. 685-689 (2016)
- [Liu 17] Liu, Q., Jiang, H., Evdokimov, A., Ling, Z.-h., Zhu, X., Wei, S. and Hu, Y.: Cause-effect knowledge acquisition and neural association model for solving a set of winograd schema problems, *Proc. 26th Int. Joint Conf. on Artificial Intelligence (IJCAI-17)*, pp. 2344-2350 (2017)
- [Luo 16] Luo, Z., Sha, Y., Zhu, K. Q. and Zhongyuan, S.-w. H.: Commonsense causal reasoning between short texts, *KR2016*, pp. 421-430 (2016)
- [Maccartney 09] Maccartney, B. and Manning, C. D.: An extended model of natural logic, *Proc. 8th Int. Conf. on Computational Semantics IWCS'09*, No. Jan., pp. 1-12 (2009)
- [Marelli 14] Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R. and Zamparelli, R.: A SICK cure for the evaluation of compositional distributional semantic models, *Lrec*, No. May, pp. 216-223 (2014)
- [Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed representations of words and phrases and their compositionality, *Advances in Neural Information Processing Systems*, pp. 3111-3119 (2013)
- [Modi 14] Modi, A. and Titov, I.: Inducing neural models of script knowledge, in *CoNLL-2014*, p. 49 (2014)
- [Mostafazadeh 16] Mostafazadeh, N., Chambers, N., He, X., Parikh, D., Batra, D., Vanderwende, L., Kohli, P. and Allen, J.: A corpus and evaluation framework for deeper understanding of commonsense stories, arXiv preprint arXiv:1604.01696 (2016)
- [岡崎 16] 岡崎直観: 言語処理における分散表現学習のフロンティア (〈特集〉ニューラルネットワーク研究のフロンティア), *人工知能*, Vol. 31, No. 2, pp. 189-201 (2016)
- [Ovchinnikova 11] Ovchinnikova, E., Montazeri, N., Alexandrov, T., Hobbs, J. R., Mccord, M. C. and Mulkar-mehta, R.: Abductive reasoning with a large knowledge base for discourse processing, *Proc. 9th Int. Conf. on Computational Semantics*, pp. 225-234 (2011)
- [Pantel 06] Pantel, P. and Pennacchiotti, M.: Espresso: Leveraging generic patterns for automatically harvesting semantic relations patrick, *Proc. 21st Int. Conf. on Computational Linguistics and the 44th Annual Meeting of the*

- ACL-ACL'06, No. Hindle 1990, pp. 113-120 (2006)
- [Parikh 16] Parikh, A. P., Täckström, O., Das, D. and Uszkoreit, J.: A Decomposable attention model for natural language inference, *Proc. 2016 Conf. on Empirical Methods in Natural Language Processing (EMNLP-16)*, pp. 2249-2255 (2016)
- [Pichotta 16a] Pichotta, K. and Mooney, R. J.: Learning statistical scripts with lstm recurrent neural networks, *AAAI 2016*, pp. 2800-2806 (2016)
- [Pichotta 16b] Pichotta, K. and Mooney, R. J.: Using sentence-level LSTM language models for script inference, *Arxiv*, pp. 279-289 (2016)
- [Raina 05] Raina, R., Ng, A. Y. and Manning, C. D.: Robust textual inference via learning and abductive reasoning, in *AAAI 2005*, pp. 1099-1105 (2005)
- [Rocktäschel 17] Rocktäschel, T. and Riedel, S.: End-to-end differentiable proving, *Advances in Neural Information Processing Systems 30* (2017)
- [Roemmele 11] Roemmele, M., Bejan, C. A. and Gordon, A. S.: Choice of plausible alternatives: an evaluation of commonsense causal reasoning, *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning* (2011)
- [Roemmele 17] Roemmele, M., Kobayashi, S., Inoue, N. and Gordon, A. M.: An RNN-based binary classifier for the story cloze test, *LSDSem 2017*, p. 74 (2017)
- [Schank 75] Schank, R. C. and Abelson, R. P.: *Scripts, Plans, and Knowledge*, Yale University New Haven, CT (1975)
- [Schoenmackers 11] Schoenmackers, S.: *Inference Over the Web*, Washington University (2011)
- [Sharp 16] Sharp, R., Surdeanu, M., Jansen, P., Clark, P. and Hammond, M.: Creating causal embeddings for question answering with minimal supervision, Technical report, *Creating Causal Embeddings for Question Answering with Minimal Supervision, Proc. 2016 Conf. on Empirical Methods in Natural Language Processing*, pp. 138-148 (2016)
- [Shibata 11] Shibata, T. and Kurohashi, S.: Acquiring strongly related events using predicate-argument co-occurring statistics and case frames, *Proc. 5th Int. Joint Conf. on Natural Language Processing*, pp. 1028-1036 (2011)
- [Snow 05] Snow, R., Jurafsky, D. and Ng, A. Y.: Learning syntactic attrens for automatic hypernym discovery, in Saul, L. K., Weiss, Y. and Bottou, L., eds., *Advances in Neural Information Processing Systems*, Vol. 17, pp. 1297-1304, MIT Press (2005)
- [Speer 12] Speer, R. and Havasi, C.: Representing general relational knowledge in conceptNet 5, *Proc. 8th Int. Conf. on Language Resources and Evaluation (LREC'12)*, pp. 3679-3686 (2012)
- [Sugawara 17] Sugawara, S., Kido, Y., Yokono, H. and Aizawa, A.: Evaluation metrics for machine reading comprehension: prerequisite skills and readability, *Proc. 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, No. 2016, pp. 806-817 (2017)
- [Sutskever 14] Sutskever, I., Vinyals, O. and Le, Q. V.: Sequence to sequence learning with neural networks, *Advances in Neural Information Processing Systems*, pp. 3104-3112 (2014)
- [Tian 14] Tian, R., Miyao, Y. and Matsuzaki, T.: Logical inference on dependency-based compositional semantics, *ACL*, pp. 79-89 (2014)
- [Tian 16] Tian, R., Okazaki, N. and Inui, K.: Learning semantically and additively compositional distributional representations, *Proc. 54th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, Long Papers, pp. 1277-1287 (2016)
- [Wang 17] Wang, Q., Mao, Z., Wang, B. and Guo, L.: Knowledge graph embedding: A survey of approaches and applications, *IEEE Trans. on Knowledge and Data Engineering*, Vol. 29, No. 12, pp. 2724-2743 (2017)
- [Weber 17] Weber, N., Balasubramanian, N. and Chambers, N.: Event representations with tensor-based compositions (2017)
- [Williams 18] Williams, A., Nangia, N. and Bowman, S. R.: A broad-coverage challenge corpus for sentence understanding through inference, *Arxiv, Proc. 16th Annual Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2018)*, to appear (2018)
- (2017)
- [Yates 07] Yates, A., Cafarella, M., Banko, M., Etzioni, O., Broadhead, M. and Soderland, S.: Textrunner: open information extraction on the web, *Proc. Human Language Technologies: The Annual Conf. of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp. 25-26, Association for Computational Linguistics (2007)
- [Yokoi 17] Yokoi, S., Mochihashi, D., Takahashi, R., Okazaki, N. and Inui, K.: Learning co-substructures by kernel dependence maximization, *IJCAI Int. Joint Conf. on Artificial Intelligence*, pp. 3329-3335 (2017)
- [Zhao 16] Zhao, K., Huang, L. and Ma, M.: Textual entailment with structured attentions and composition, *Proc. COLING 2016, the 26th Int. Conf. on Computational Linguistics: Technical Paper*, pp. 2248-2258 (2016)
- [Zhao 17] Zhao, S., Wang, Q., Massung, S., Qin, B., Liu, T., Wang, B. and Zhai, C.: Constructing and embedding abstract event causality networks from text snippets, *Proc. WSDM2017*, pp. 335-344 (2017)

2018年2月20日 受理

著者紹介



井之上 直也 (正会員)

2008年武蔵大学経済学部経済学科卒業。2010年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。2013年東北大学大学院情報科学研究科博士後期課程修了。2013年株式会社デンソー基礎研究所研究員。2015年より東北大学大学院情報科学研究科助教、現在に至る。博士(情報科学)。自然言語処理の研究に従事。言語処理学会、情報処理学会各

会員。