

Deep Learning と人工知能の発展

Deep Learning and Advancement of AI

松尾豊^{*1}

Yutaka Matsuo

^{*1} 東京大学

University of Tokyo

This study tries to identify the essential advancement of AI focusing on deep learning. Recently deep learning has attracted much attention. Its core contribution is to show the feasible approach to obtain “representation” automatically without human intervention. Although the technology is not mature yet, it has a huge potential because the problem of acquiring representations has been a key issue in AI. However, the theoretical consideration on representation is not conducted enough and is needed for further development of representation learning. In this paper we attempt to discuss what is representation, what consists a representation, and what is the meaning of obtaining representations. Such framework would contribute to the future discussion and improvement of representation learning.

1. はじめに

近年、Deep Learning が注目を集めている。Deep Learning は、観測データから本質的な情報を抽出した内部表現／潜在表現／特徴量を学習する。多層であることは、多くの場合に利用可能なヒューリスティックのひとつであるので、そのインパクトの大きさを語る際には、より抽象的には、表現学習 (representation learning) と置き換えたほうが適切であろう。

表現を学習するとはいったい何だろうか？この問いはシンプルであるが、実は大変に難しい。というのも、これまでに人工知能が扱ってきた問題のほとんどは、表現が決まった後のモデル化や最適化であるためである。表現を学習することを考えると、「可能な表現」という、表現の学習における解空間について考える必要がある。ところが、可能な表現を考えるには、また表現を使わなければならない。したがって、可能な表現が何であるか、それがどのように定義され得るかというのは難しい問題であると考える。

本稿では、表現学習とは何かについて述べる。表現とは何か、それを構成するものは何か、そして、表現を学習するとは何かについて順に述べる。かなり粗い議論であり、まだ不十分な思考の過程であるが、こうした試みにより、表現学習のより効率的なアルゴリズムの開発や理論面での整備につながることを期待している。

2. 表現とは

表現学習における「表現」は非常に難解な概念である。英語の representation という言葉は、represent (代表するもの) という意味合いがあり、ものごとを代表して表すもの、という意味合いがある。あるいは、re-present (再び-現れる) という意味もある。まさに Deep Learning が、自己符号化器で情報をよく復元するような表現をつくるというニュアンスが見事に含まれている。

ところが日本語で「表現」というと、たとえば、文学や絵の作品のような「表現されたもの」をイメージしてしまう。歴史的には、人工知能では knowledge representation を知識表現と訳してきたが、知識を「表現する」とこと、特徴が「表現される」ことは主体、

客体が異なり、「表現される」という意味で使うのは日本語だと少し違和感がある。哲学では「表象」と翻訳されることもある。本稿における表現とは、特徴が「表現される」ものであるということに注意いただきたい。

この場合の表現を、情報処理の点から捉える。表現の問題とは、現実世界から観測するデータをもとにして、いかにして表したいもの (オブジェクトやイベント、現象など) の特徴を表すかという問題となる。

つまり、現実世界から観測するデータ X があったときに、これを使って定義される何らかの関数 $f(X)$ が表現 (あるいは素性) である¹。 $f(X)$ は 0-1 の値を取ると考えても良いし、何らかの連続値を取ると考えても良い。ある画像を入れた時に猫が映っているかを判定する関数 $f_1(X)$ は猫の表現であり、ある動画から不審人物がいるかどうかを反転する関数 $f_2(X)$ は不審人物の表現である。

ところが、こうして考えた時に、いろいろな疑問が浮かぶ。 f はどのように定義されるのだろうか。 X はどのように定義されるのだろうか。以下では、 X が有限な集合に限定されている場合と、そうでない場合について分けて考えよう。

2.1 観測データ X が有限である場合

X が有限な場合に、そこからどのような表現を作ればいいのかは、比較的容易な問題である²。表現となる関数の解空間としては、観測データ X 上で定義される任意の関数はもちろん、時間的な広がりをもった観測データ X の過去のデータを使うこともできるし、それらの複雑な演算結果を使うこともあり得る。(その場合、通常は手続き的に定義されることになる。) そういった全ての可能性を含めた関数 f が、表現の候補となる解空間になる。

そして、問題は、この膨大な表現の解空間をどのように効率的に探索すればよいかということである。その際にはさまざまなヒューリスティック、あるいは prior がありそれらを活用して効率的に探索していくことになる。(Bengio による 10 の prior などを参照されたい [Bengio 13].) 効率性というのが現実には最も大きな

¹ある表現が、機械学習のための特徴量として用いるときは素性と呼ばれるため、本稿では、表現と素性という用語は、相互交換可能なものとして扱う。

²容易であるといっても相対的な話であって、そもそも極めて難しい問題であることには違いはない。

問題であり、原理的に探索可能であるということと、現実的な計算量で探索可能であることが異なることは周知の事実であることは改めて言うまでもない。Deep Learning は、(ある限定された)表現の空間を、効率的に(といっても大変な計算ではあるが)探索できるところが優れた点である。

2.2 観測データ X が決まっていない場合

問題は、観測データ X が有限な集合として定義されていない場合である。これは言い換えれば、「世界のどの情報をどのような方法で取得しても自由であり、そうしたありとあらゆる情報の取得の仕方を許容したときに、そのとり得る情報の全体集合を定義することができるか」ということでもある。

世界からとり得る情報は無数にある。例えば、ある地点のカメラにどんな画像が映ったか。ある人があるお店でどの商品を買ったか。ある人工知能学会の論文原稿にどのような単語が含まれているか。人工知能学会の論文原稿が、締め切りに間に合ったのか間に合わなかったのか。こうした情報が無数にあるのはまだ良いとして、問題はこの「情報」の定義に高度な認知的な処理が必然的に含まれてしまっていることである。

ある「人工知能学会の論文原稿」といったときには、ひとつの論文原稿を認識するという認知のプロセスが含まれている。つまり、ある論文原稿がいまここにあって、明日は別の場所にあるともそれらは同一であり、それが電子的にアップロードされていると同じであるということが了解されている。お店で買う商品もそうであり、お店の同一性や商品の同一性が暗黙に了解されている。ある地点のカメラ映像ですらそうであり、例えば、自転し公転する地球上で、地球上の同一地点という同一性の認識によってはじめて、ある「地点」のカメラの映像が定義される。そして同一性の認識には、あることとあることが同一であることを確かめる処理が必要であり、特徴を抜き出す、つまり表現を生成する必要がある。

すなわち、表現の候補となる解空間を定義するために X を定義する必要があるわけだが、その X の定義自体に同一性の認識が必要であり、表現を使う必要がでてしまう。これでは表現の候補となる解空間を定義できたとはいえない。

2.3 原始センサ

では、表現に依存しない X の取り方はないのだろうか。同一性の問題を回避して定義できるのは、唯一、「あるセンサが観測した値」という定義の仕方だけであろう。つまり、「何かの値」を観測するのではなく、「観測してしまった値」から話をスタートするしかないということである¹。

この場合でも「あるセンサの同一性」をどう考えるのかという問題があるが、ここではやむを得ずセンサの永続性は仮定するものとする(センサの永続性まで担保できないとするとその後の議論ができないので)。そうすると、あるセンサが時系列でとった値を用いることができる。時間を特別扱いする必要のあるのかも疑問であるが、ひとまずここではセンサの永続性および時系列の値が使えることとしよう。

「観測してしまった値」からスタートするとして、通常、センサがどのように実現されるかを考えると、そこには非常に大きな文脈依存性がある。例えば、海中の水温を図るためのセンサを港につければ、港の水温が計測されるが、それが船底に設置されていれば、船がいる場所の水温を図ることになる。また、鼓膜は空

気の振動の圧力を感じることで音を検知しているが、デバイスの形状によって取りうるものも大きくことなる。

ここでは、「センサの同一性」と「そこで時間的に連続した値が計測される」ということのみを仮定し、その値の計測に何らの同一性の解釈(あるいは表現)を伴わないようなものを、ここでは原始センサと呼ぶことにしよう。原始センサは、本質的に非常に高い文脈依存性を持つが、それについては感知しない。

3. 無限の素性空間

こうしていったん原始センサが定義できると、可能な表現(あるいは素性)は、原始センサの値に対する関数として定義することができる。あらゆる原始センサの値を用いた任意の演算や操作の組み合わせが、新たな値を生成し、それがありとあらゆる素性の空間を構成する。この素性には、表現による同一性を内包するものも含まれる。したがって、「ある地点の温度」や「ある人がお店で何を買ったか」なども素性のひとつである。そして、この世界は、時々刻々と変化する無限の素性で満ちた空間であると言いつてもできる。

さて、この無限の素性空間の性質はどのようなものだろうか。まず、ある素性が他の素性に対して情報量を持つことがあり得る。例えば、ある原始センサに 1 を足した素性と、そのセンサに 2 を足した素性は、完全に相関する。また、ある原始センサの「近く」に設置された原始センサの値は、相関が高いだろう。このように自明な相関のほかにも、雨が降ると池の水は増えるだろうし、エアコンをつけると部屋の温度が変わるだろう。このように、素性間の情報量を定義することができる。

さらに時間の先行性を考慮し、予測性の観点から素性の関係を定義することもできる。ある素性の組み合わせが、他の素性に対して予測性を持つのであれば、それらの間には関係が定義できる。すべての素性は、原始センサの関数であるとしているので、その関数の形によっては、ある素性がある素性の関数として完全に記述できる(あるいは予測し得る)こともあり得る。

4. 主体と予測性、世界への働きかけ

世界をこうして原始センサ、およびその上に定義される素性空間から成るとしたときに、議論を進める上で、次に導入しなければならないのが、「主体」である。素性における情報量や予測性を活用するための主体を導入する。

こうした情報空間における主体の定義も甚だ困難であるが、ここでは、自己同一性を保持・保存したいオブジェクト²を主体と考える。主体にとっての目的は、自己同一性の保持・保存であり、生物にとってそれは個体や種の維持・保存である。

主体にとっての目的にどのくらい資するかという観点から、ある素性の効用を定義することができる。ある素性が利用可能になると、世界の予測性が上がり、主体の行動が洗練されることなどによって自己同一性の保持・保存の確率が上がる。

この主体が利用可能な原始センサがある場合に、この主体にとって都合のよい情報処理とは、原始センサを使って構成可能な素性のうちで、効用の最も高い素性を発見し活用することである。通常は、原始センサからの有用な素性の発見は山登り法的方法で行われる。Deep Learning で採用されている方法は、自己符号化器によって、層ごとに素性を構成していく方法である。山登り法なので、必ずしも最適解が得られているとは限らない。ある素性の組み合わせがどのような効用を持つかは、現実世界において特定の構造が機能の発現するさまと同じく、著しく

¹ この場合の観測とは何か、観測する主体は何かなどの問題があるが、ひとまず置いておく。

² オブジェクトと言っても、ここにも同一性の認識が内包してしまう(泣)。

非線形であり、いくつかのヒューリスティックに頼って探索する他ないと思われる。

主体にとって、原始センサの設計自体も探索の範囲である。したがって、センサの形状を変えたり性質を変えたりすることによって、原始センサも変化しうる。(生物の場合には進化によって変化する。)

また、主体が世界に対して働きかける方法が増えると、同じ原始センサでも異なる情報をとり得る。例えば、温度計でも、主体が地中深く掘ることができる能力があると、地中深くの温度が取れるし、動物を捕まえることができると動物の温度を測ることもできる。主体の世界への働きかけの方法は、通常は有限の手続きを組み合わせることによって行われるが、主体の世界への働きかけ自体によって、原始センサで取れる情報が異なってくるということは、異なる種類の原始センサになる、つまり原始センサの種類が増えるということである。

主体の世界への働きかけは、同一性の認識を含んだ表現の問題を内包するから、主体の世界への働きかけを考慮すると、結局は、表現と世界の認識というのはループになってしまうことは避けられない。かくして、主体のもつ原始センサ→表現(素性)→主体の世界への働きかけ→主体のもつ原始センサというループを繰り返しながら、自己同一性の保持・保存に資するための原始センサと表現の獲得、および世界への働きかけの方法に関しての解空間を探索しているということになる。

つまり、表現学習の問題において、可能な表現の解空間を書き表すには、1) 原始センサの無限性、2) 関数の無限性、3) 現実世界の働きかけの無限性という3つの無限性に対処しなければならない。このことを考えると、可能な表現の解空間を仮定することは難しい。現実的には、原始センサを固定すること、関数を制限すること、働きかけの種類(可能な行為)を制限すること等によって対処するしかないだろう。(しかし、この無限性を考慮した上で、適切な絞り方をする必要があるだろう。)この解空間を少ないサンプル数からできるだけ絞るための工夫としては、Deep Learning においては、ノイズを加える、Dropout を使うなどであり、また人間が利用している言語の活用、社会を構成することなどもこうした方法のひとつであると考えることができる。

5. まとめ

本稿では、世界を満たす無限の素性、それを構成するための原始センサの概念、およびその素性空間上で得られた情報量を活用するための主体の概念について説明した。本稿で言いたいことを箇条書きでまとめると以下である。

- ・ 世界は素性で満ちている。
- ・ 定義上、同一性の認識を含まないセンサを原始センサと呼ぶ。素性は原始センサの関数(あるいは手続き)として構成される。
- ・ ある素性は、ある素性に対して情報量を持つ。また時間的に先行する予測性を持つ。
- ・ これを利用することにより、ある素性を主体の目的に利用することができる。
- ・ 可能な素性(あるいは表現)の空間は、i) センサの設計、ii) 値の加工、iii) 主体の世界への働きかけ の 3 つから構成される。
- ・ 素性の組み合わせは、実世界における特定の構造が機能を発現するのと同じ程度に非線形で複雑である。
- ・ この探索のためには、少ないデータから解空間をいかに効率的に絞り込むかが鍵であり、現実的な解法として、通常は山登り的な方法によりアドレスされる。また、

Dropout、ノイズの重畳、複数のエージェント、言語などもそのための方法と考へ得る。

最後に、Deep Learning をはじめとする表現学習を理論的に扱い、さらに効率的なアルゴリズムの構築を目指すには、表現学習の問題を適切に定義する必要があると感じている。表現の問題は難しく、いまのところ表現を探索するという点に関しての理論的な枠組みはほとんど整備されていないのではないだろうか。過去の試みも踏まえつつ、Deep Learning が現実的に実現できるようになった現在において、そうした理論的な枠組みの整備を進めていく必要があるのではないかと感じている。

参考文献

- [Bengio 13] Y. Bengio and A. Courville and P. Vincent: Representation Learning: A Review and New Perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 8, 2013
- [松尾 14] 松尾 豊: AI の未解決問題と Deep Learning, 人工知能学会全国大会, 3H-OS24b, 2014.