

機械学習: もう一つの選択

佐藤理史 京都大学工学部

sato@kuee.kyoto-u.junet

1 はじめに

機械学習の手本は、人間の学習である。人間の行なっている学習と同等なものを計算機上で実現するというのもって機械学習とするならば、「機械学習は可能である」と筆者は考える。その根拠は、特に不可能だと考えるべき積極的な理由が見つからないからである。

学習は魔術でもなければ奇跡でもない。必ず種があるし、種明しされてしまえば、なんだと思うようなものであるはずである。ある意味では、学習の研究は、種として何をもってこればよいか(人間はどんな種をもっているのか)、あるいは、現実的に用意できる種から最大限の効果を引き出すにはどうしたらよいかということを考えるということになる。

しかしながら、機械学習をとりまく環境は、それほど単純ではない。アルゴリズムがわかっている問題を解くことが人工知能に課せられると同じように、人工知能の中で難しい問題を解くことが機械学習に課せられているように見受けられる。まるで人間ができないことを、機械が学習能力を持つことによって一挙に解決するかのような期待感がそこはかたく感じられる。しかし、それは幻想であり、機械学習に対してフェアではない。

本稿では、筆者が考える機械学習の当面の目標とすべき対象と、それを進めるときの2つの仮定について述べ、そこから導かれる「規則の一般化」への懐疑と、それに対するもう一つの選択としての「記憶を中心とした学習」について述べる。

2 2つの仮定

我々は機械学習に何を求めているのだろうか。人間が未だかつて解けなかった問題を解くような

「天才」を求めているのか。膨大な言語データから、完備な言語理論を構築することを求めているのか。

勿論、それが出来れば素晴らしいことであるし、ある意味では人間を越えたことになるかもしれない。しかし、我々は、まず、普通の人間が誰でも持っている普通の学習能力に近づこうと努力すべきである。すなわち、筆者の考える機械学習の当座の目標は、普通の人間が日常的におこなう、ちょっとした知識の拡大である。例えば、チーズをはさんでポークチーズかつを作る方法は、チキンにも応用できるはずだとか、バドミントンでドロップショットが決まったならば、もう一度試してみるとか、そのような学習である。

このような学習のモデル化、及び、実現を考えたいく前提として、筆者は以下の2つを仮定する。

- ほとんどの学習は、単純なメカニズムでなされている(実現できる)に違いない。すなわち、学習アルゴリズムは単純であればあるほどよい。
- 学習(系)は、実行(系)と切り放して考えることはできない。多くの学習は、実行(あるいはリハーサル)時に起こる。

この仮定には、単なる仮定ではあるが、このような学習は通常意識されないと考えられるので、

- 意識的な処理に相当するような、複雑な記号処理的プロセスが動いているとは考えにくい
- なんらかのトリガー、あるいは、要求がないかぎり、知識の拡大がなされるとは考えにくい

ということを念頭においている。

3 「規則の一般化」への懐疑

近年の機械学習の主流は、「規則の一般化」という言葉で総括することができるだろう。すなわち、明示的な規則表現を採用し、いくつかの実例から規則を生成する。そこでの中心的な考え方が、規則を一般化するという考え方である。

しかし、前節の仮定に立って考えるならば、この方向は何か少し違うのではないかと思われる。主な論拠は、

- 学習系と実行系が独立である。(実行に先立ち、すべての実例空間を線引きする必要が本当にあるのか。)
- 一度学習された結果を修正することが容易ではない。すなわち、知識表現が学習に対して柔らかくない。(知識の体系は、学習に対してもっと柔軟なはずである。)
- 学習アルゴリズムが複雑である。
- 人間の単純な知識の拡大というよりは、データから理論を構築するモデルに相当しているのではないか。

である。

4 記憶を中心とした学習

確かに、「規則の一般化」という考え方は、能力の向上ということを見ることができるといって、受け入れやすい。しかし、明示的ということを除いて、単に実行能力の拡大ということを考えるならば、もう一つの選択肢が存在する。

既に、筆者は、[Sato 89b]において、Memory-based Translation(MBT)における実行能力の拡大と、「規則の一般化」の関係を論じ、その対応点と相違点を明らかにすることを試みた。その結果、MBTは、明示的な規則の一般化なしに、それと同等の能力の拡大を実行時に行なっていることが明らかになった。このことは、「規則の一般化」を中心としない、もう一つの方法、すなわち、記憶を中心とした学習の可能性を示唆する。

記憶を中心とした学習は、暗記学習を基本とする。すなわち、実例を記憶し、それをうまく利用

することによって、規則の一般化と同等の効果をあげることをもくろむ。その中心的な課題は、

1. 記憶の組織化の問題(与えられた実例をどのように構造化して記憶するか、どのように選択的に実例を記憶するか)
2. 記憶の利用の問題(記憶した実例をどのようにして利用するか、特に、類似性の利用や best match の手法)

である。大まかな分担としては、1.が抽象化を受け持ち、2.が暗黙的な一般化(適用範囲の拡大)を受け持つ。Memory-based Translation[Sato 89a]は、その最も単純なモデルの1つと考えることができる。

記憶の組織化で用いられる構造は、当面は与えられたものとして考えるが、より野心的な目標としては、この構造を変更していくということも考えられる。フレーム問題の疑似解決問題 [Matsubara 89]は、記憶の組織化の変更ということで捉えられかもしれない。

5 おわりに

本稿では、「規則の一般化」に代わるもう一つの選択肢として、記憶を中心とした学習の可能性がありうることを述べた。筆者が進めている Memory-based Translation Project では、機械翻訳という領域において、記憶を中心とした学習を考えると同時に、他の学習方法との関連、及び、人間の学習との関連等も併せて考察していきたいと考えている。

参考文献

- [Matsubara 89] 松原仁, フレーム問題に対する考え方, Workshop on Learning 89, 1989.
- [Sato 89a] 佐藤理史, 長尾真, 実例に基づいた翻訳, 情報処理学会研究会報告, NL-70-9, 1989.
- [Sato 89b] 佐藤理史, *Memory-based Translation* と規則学習, Workshop on Learning 89, 1989.