

# 自然言語処理分野における機械学習研究の位置づけ

鈴木 潤

NTT コミュニケーション科学基礎研究所

suzuki.jun@lab.ntt.co.jp

## 概要

構文解析，機械翻訳といった最も歴史ある言語処理分野の二大研究テーマは，近年では機械学習を含む統計的なアプローチが主流になっている．また，機械学習手法は，それ以外の様々な自然言語処理タスクを解決する際にも，直接または間接的に取り入れられ，多くの自然言語処理タスクの発展に貢献してきたと言える．

しかし，このような自然言語処理研究における成功の弊害として，機械学習手法が拡大解釈され，使い方を誤った研究がなされている場合が見受けられる．端的な例として，機械学習手法の成功は，まず対象とする問題全体を人間がルール等で網羅できない程度以上には複雑な問題であることが大前提である．次に，そのようなある程度以上複雑な問題において，統計量が十分に計算できる程度のデータ量が存在することが挙げられる．つまり，人間が簡単にルールで網羅できるような簡単な問題であれば，敢えて機械学習を持ち出す意味はあまりないし，利用可能なデータ量が解きたい問題に対して非常に少なければ，どんなに良い機械学習手法を用いても，一般的には良い結果は得られない．しかし，前述のような前提を満たさない問題に対しても，盲目的に機械学習手法を試みるような研究も行われる場合もある．これ以外にも，自然言語処理分野では機械学習の誤った拡大解釈や利用がしばしば起こっているように思われる．

そこで本発表では，自然言語処理分野における機械学習研究の立場から，いくつかの研究成果を事例として，主に自然言語処理分野における機械学習手法の解釈を整理し，その位置づけを明確にする．それによって，自然言語処理分野における機械学習の誤った解釈を正すことを試みる．また，自然言語処理分野における機械学習研究の今後の展望について意見を述べる．