



普遍論理としてのTransformer理解

2023年8月8日

[情報統合技術研究合同会社](#)

岡島 義憲

okajima@info-integnology.com

自己紹介：岡島義憲

■ 経歴

- ・ 1981年～2019年 富士通半導体設計部門
Bipolar Memory、SRAM、DRAM、Micro Controller、ASIC、ASSP
2015年～2019年は、SIRIJ、JEITA、SEAJ等、経産省と半導体業界団体
- ・ 2021年4月 情報統合技術研究合同会社を設立
(AI回路の知財開発、ロジック系集積回路設計への問題提起)
(<https://info-integnology.com/index.html>)
- ・ 2015年から、汎用AI開発に興味を持つ。
 - 今は、AI回路をGPUを使って回路シミュレーションしている段階だが、設計が完了すると、次は、その回路を搭載したチップ開発の時代が来る。
 - 生成AI や GPTは、想定外の進展
 - Transformerは、AI回路を構成する普遍論理(Universal Logic) だろう。

Transformer アーキテクチャ

~~ そのAGI的な推論能力がどこから来ているのか？ ~~

- 構成 : 1) Sequence To Sequence変換 ベクトル列の変換
2) Attention機構の導入 単語推測 & 構文推測
3) 非再帰構造 / Residual接続
4) 大規模化(パラメータ大 / 学習データ大)
5) One-Shot (Prompt) 学習
- } 学習能力

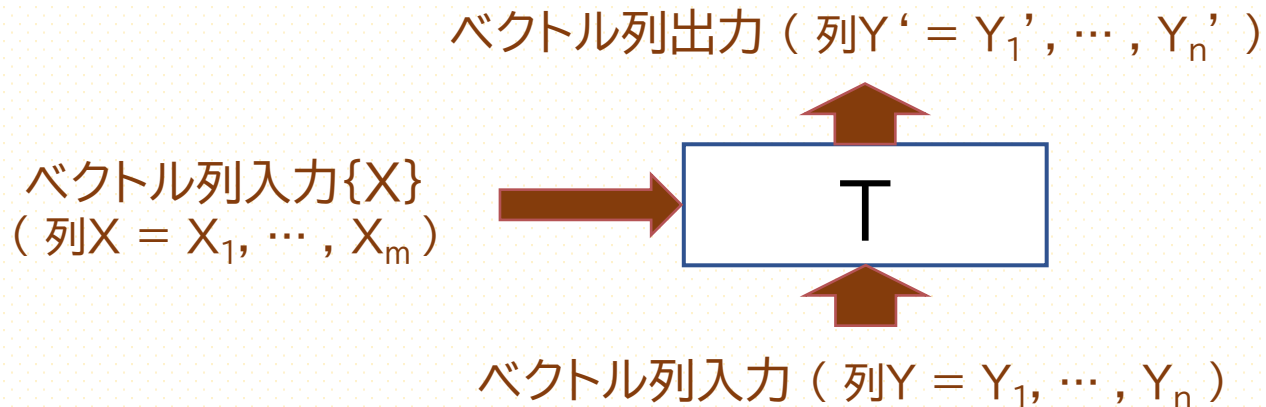
「Attention機構」は、普遍的(Universal)な推論論理か？

入力情報を元に、出力側の「単語(Symbol)と構文(Pattern)」を同時最適化
(相互依存の大きい2種の確率分布)

T回路

T回路モデル：ベクトル列を変換する回路

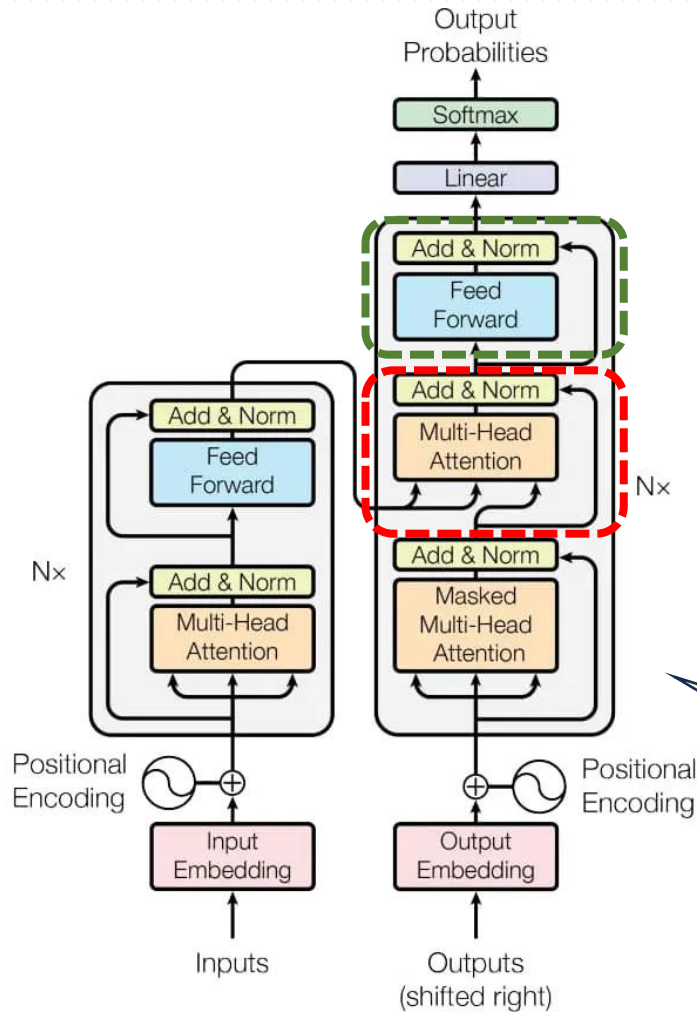
1. 入出力モデル



2. 基本機能

左からのベクトル列入力(X)によって、
下からのベクトル列入力(Y)を確率的に変形(変換)させて、
上辺に、ベクトル列(Y')出力する回路

Google社のTransformer



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$c_i = \sum_j a_{ij} h_j \quad (1)$$

$$a_{ij} = \frac{\hat{a}_{ij}}{\sum_j \hat{a}_{ij}} \quad (2)$$

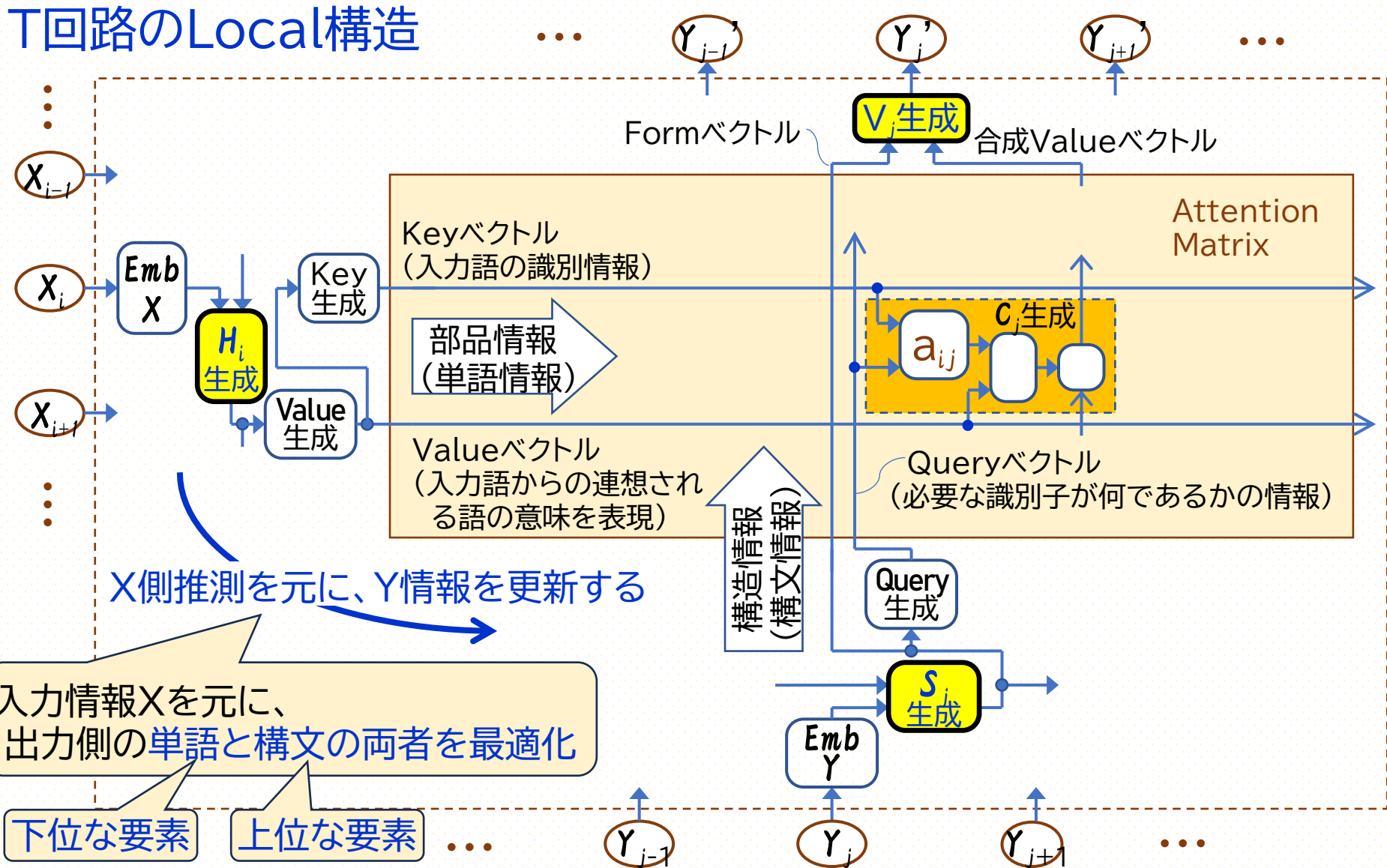
$$\hat{a}_{ij} = \text{att}(s_i, h_j) \quad (3)$$

$$\text{att}(s_i, h_j) = \langle W_h h_j, W_s s_i \rangle.$$

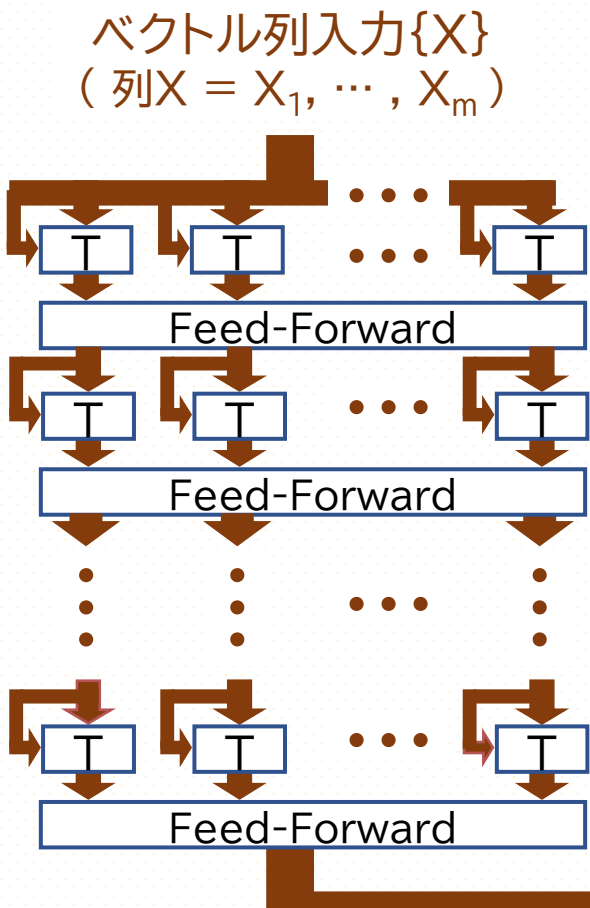
$$P(y_i | y_1, \dots, y_{i-1}, \mathbf{x}) = \text{softmax}(W[s_i; c_i] + b)$$

Attention機構は、知的情報処理に普遍的に現れる回路構造といえるか？

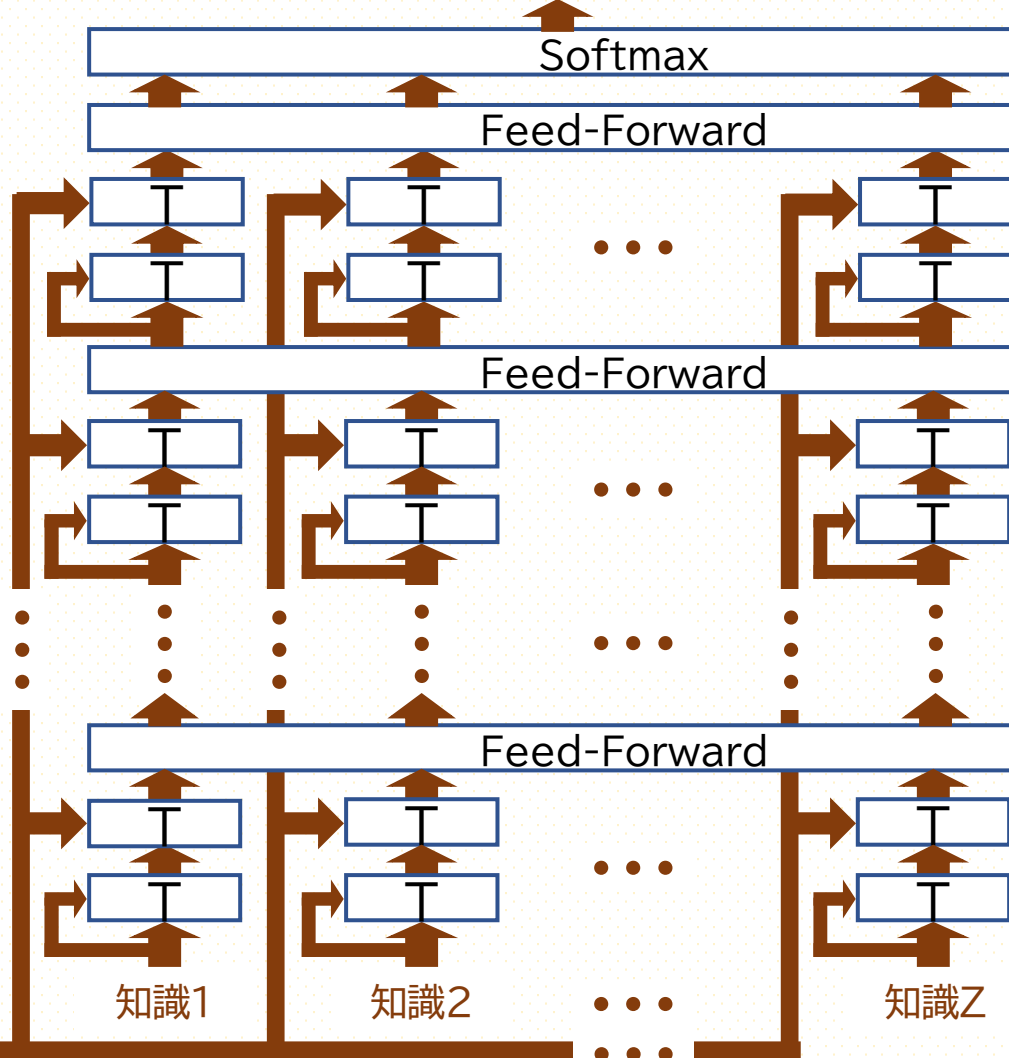
T回路のLocal構造



T回路の展開 (Transformerの構成)

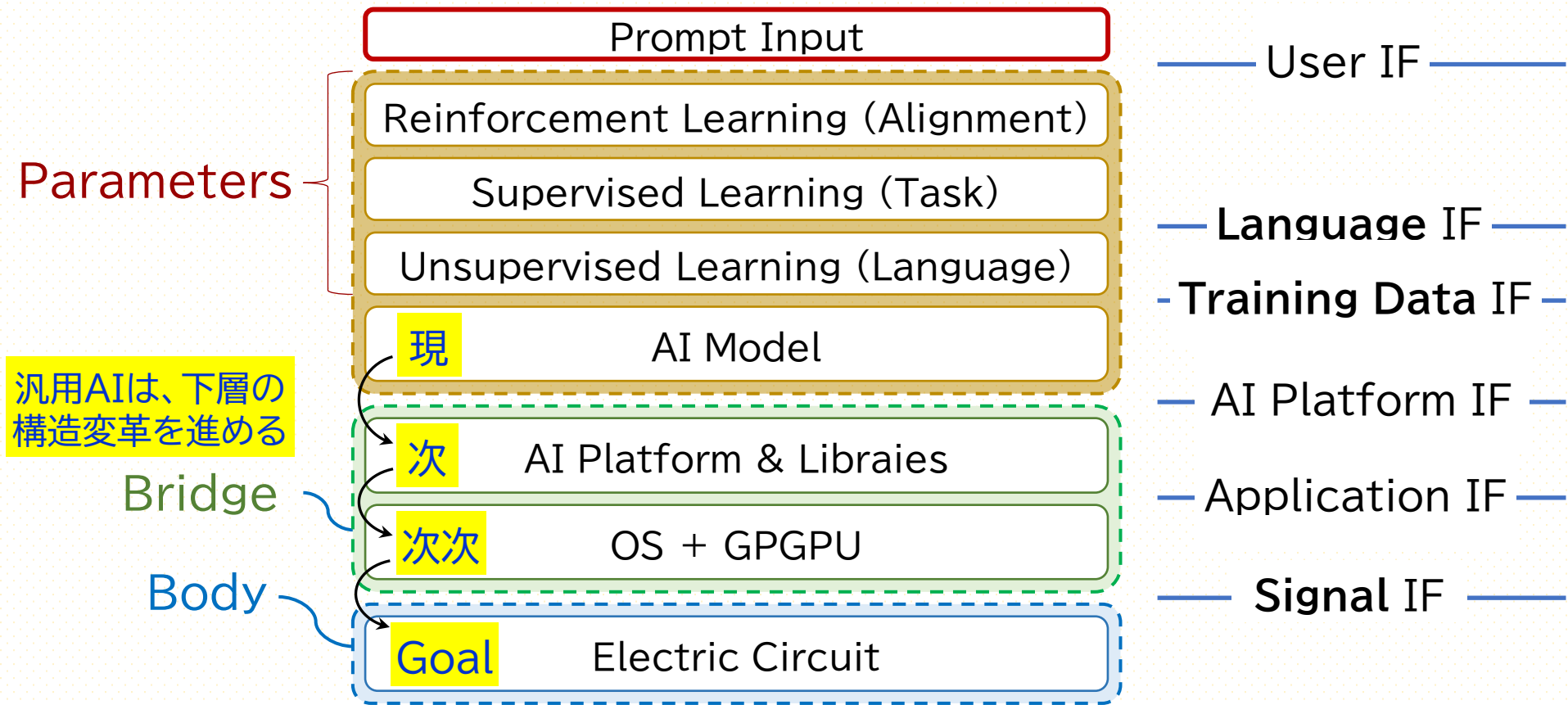


ベクトル列出力 (列Y' = Y'_1, \dots, Y'_n)




AI機能の階層構造上のT回路の実装位置

汎用的(普遍的)な機能は、より下層にてモジュール化した方が「性能」と「効率」が良い筈



まとめ

- Transformerアーキテクチャを電子回路に見立てて、T回路を用いて表現してみました。
- T回路が知的コンテンツ生成で汎用的な論理であるとする、それは、相互依存の大きい2種の情報（SymbolとPattern）の最適化を行うというプロセスが普遍的に存在するからだろうと見做した。
- T回路のような汎用的/普遍的(Universal)な論理は、いずれ、
Platform化 ⇒ OSへの繰り込み ⇒ ハードウェア化
するだろう予想する。（下層で用意した方が、効率や性能が高くなる）



ご清聴いただき、ありがとうございました。