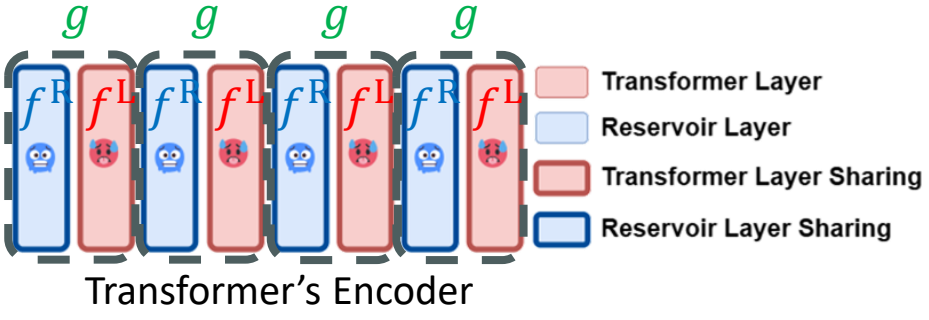
2024年度未踏ターゲット事業（リザバーコンピューティング技術を活用したソフトウェア開発分野）

リザバーコンピューティングを用いた次世代型自然言語処理モデルの実現
― リザバー計算に触発された軽量型Transformer の提案 ―

中村 仁（大阪大学 大学院情報科学研究科）

【背景・目的】
- 近年の言語モデルは大規模化が進み、学習・推論に膨大な計算資源を要する
- Transformerモデルの軽量化に関する研究が進展している
- 軽量な計算を行うリザバーコンピューティングの仕組みをTransformerに取り入れ、省メモリかつ高速な自然言語処理モデルの構築を目指す



【開発したソフトウェアの特徴（新規性・優位性）】
- Transformerの一部パラメタを共有・固定化し、学習コストを削減しながら精度を維持
- ハードウェア実装を意識した設計により、リソースを最適化した推論を実現

【解決する課題と社会への影響】
- 演算量と電力消費を低減し、オフライン環境やエッジデバイスにおける高精度な言語処理が期待される
- 機密情報の保護や災害現場等、インフラ制約のある状況でも活用が期待される

【開発したソフトウェアのアピールポイント】
- 計算負荷を抑え、短時間で収束する軽量設計
- FPGA実装を想定した設計で多様な応用領域へ展開可能