



AAMT
Asia-Pacific Association for Machine Translation
アジア太平洋機械翻訳協会

機械翻訳の最新技術動向調査 ～特許翻訳への応用を見据えて：経過報告～

須藤 克仁

AAMT/Japio特許翻訳研究会 副委員長
(奈良先端科学技術大学院大学)

自己紹介

- 須藤 克仁（すどう・かつひと）
 - AAMT/Japio特許翻訳研究会 副委員長
 - 奈良先端科学技術大学院大学 准教授
 - （兼）科学技術振興機構 さきがけ研究者
 - （兼）理化学研究所 革新知能統合研究センター 客員研究員
 - 機械翻訳・音声言語処理の研究に従事



AAMT/Japio特許翻訳研究会

- 日本特許情報機構(Japio)の委託で2003年発足
 - 特許翻訳に関する研究活動
 - シンポジウム・国際ワークショップの開催
- 構成員
 - 委員長: 辻井 潤一
 - 副委員長: 宇津呂 武仁, 須藤 克仁
 - 委員: 江原 暉将, 黒橋 禎夫, 越前谷 博, 二宮 崇, 綱川 隆司, 後藤 功雄, 今村 賢治, 中澤 敏明



AAMT
Asia-Pacific Association for Machine Translation
アジア太平洋機械翻訳協会

【2020年度の活動】 機械翻訳の技術動向調査

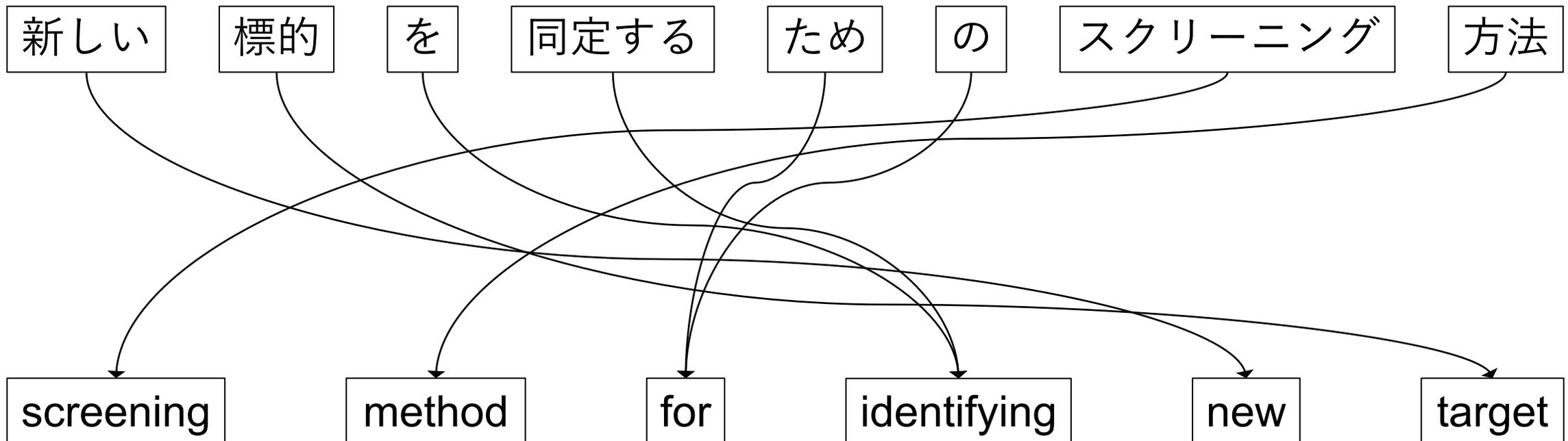
技術動向調査の背景

背景: 急速な技術の変化

- 1950年代から: **規則ベース**機械翻訳
- 1980年代初頭: **用例ベース**機械翻訳 (長尾先生)
- 2000年頃から**統計的**機械翻訳の研究が加速
 - 当初の試みは1980年代後半
 - 2005年 [米NISTの共通タスク](#)でGoogleが他を圧倒
 - 2006年 Google Translate 公開
- 2010年代半ば: **ニューラル**機械翻訳

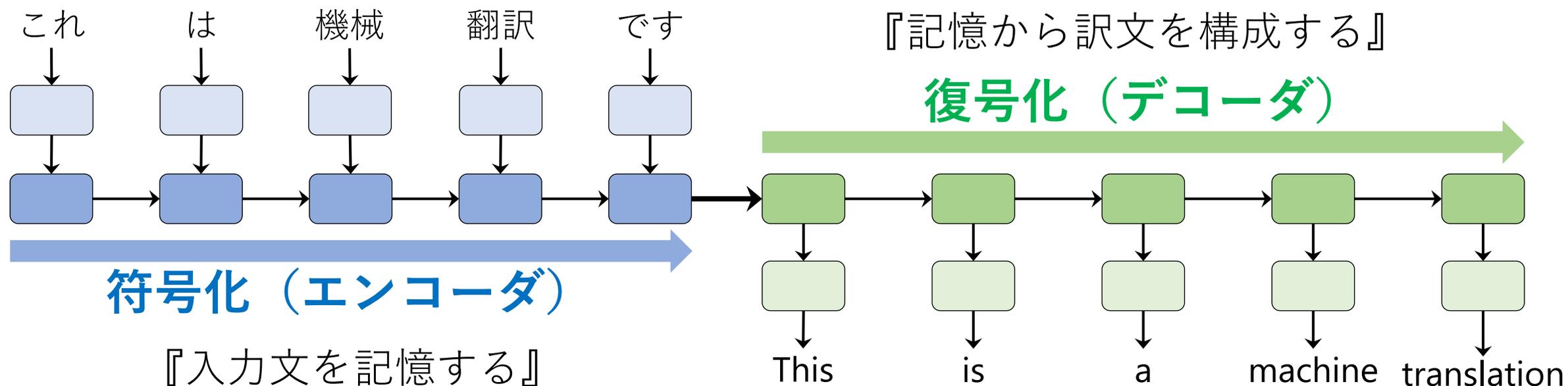
ニューラル機械翻訳以前の機械翻訳

- 構成性 (compositionality) に基づく機械翻訳



ニューラル機械翻訳

- 要素毎の書き換えから脱却した流暢な訳出
 - 構成性が見かけ上満たされないことも



翻訳プロセスの考え方の相違



出典: デイリーポータルZ
<https://dailyportalz.jp/b/2006/04/04/c/2.htm>



出典: THE VERGE
<https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan>
 (thispersondoesnotexist.comで生成)

【従来の考え方】

部品を揃えて組み合わせる

【NMTの考え方】

「記憶」を辿ってそれらしく生成

プロセスの違いによる差①



【従来型機械翻訳】

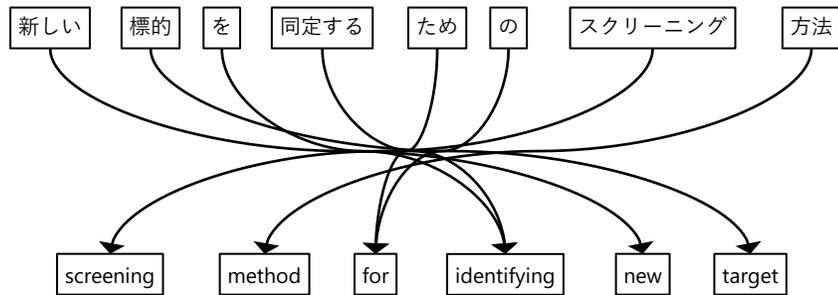
- 部分部分はそこそこ正しい
- 生成文が流暢でないことも
- 木を見て、森を見ず



【ニューラル機械翻訳】

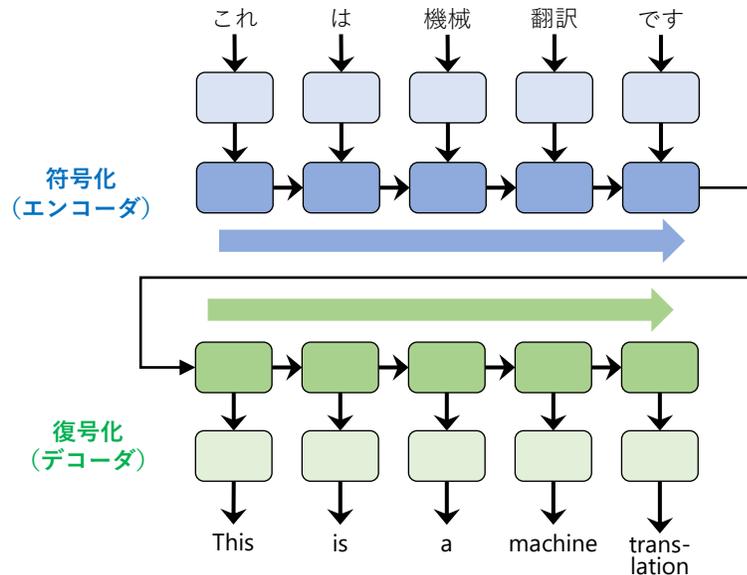
- 条件が整うと非常に強力
- 信じられないような誤訳も
- 森を見て、木を見ず

プロセスの違いによる差②



【従来型機械翻訳】

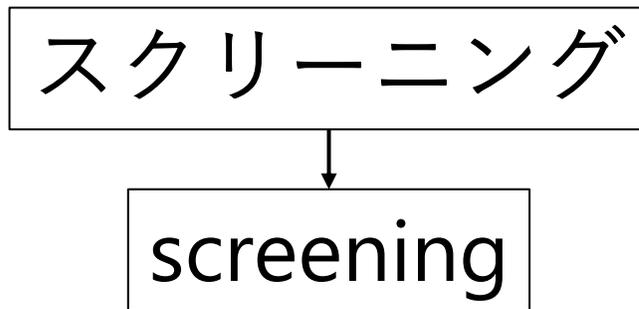
- 入力を全被覆
- 抜けや重複が少ない



【ニューラル機械翻訳】

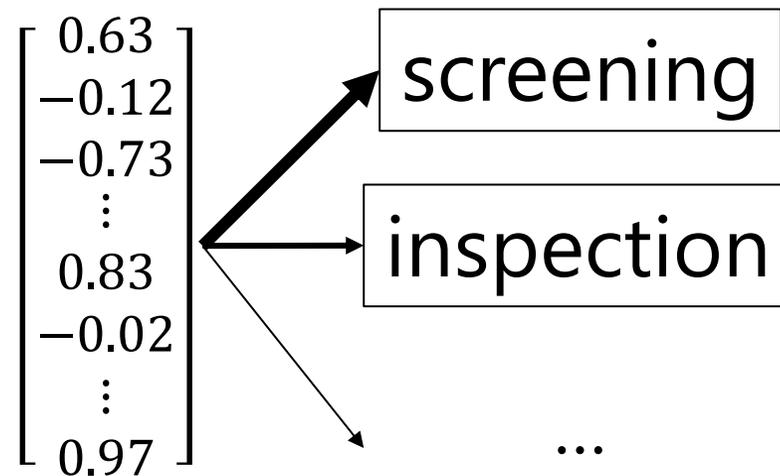
- 入力の全被覆は不保証
- 抜けや重複の可能性も
- そもそも「変換」か？

プロセスの違いによる差③



【従来型機械翻訳】

- 記号の対応が明確
- 記号間の類似性は未知

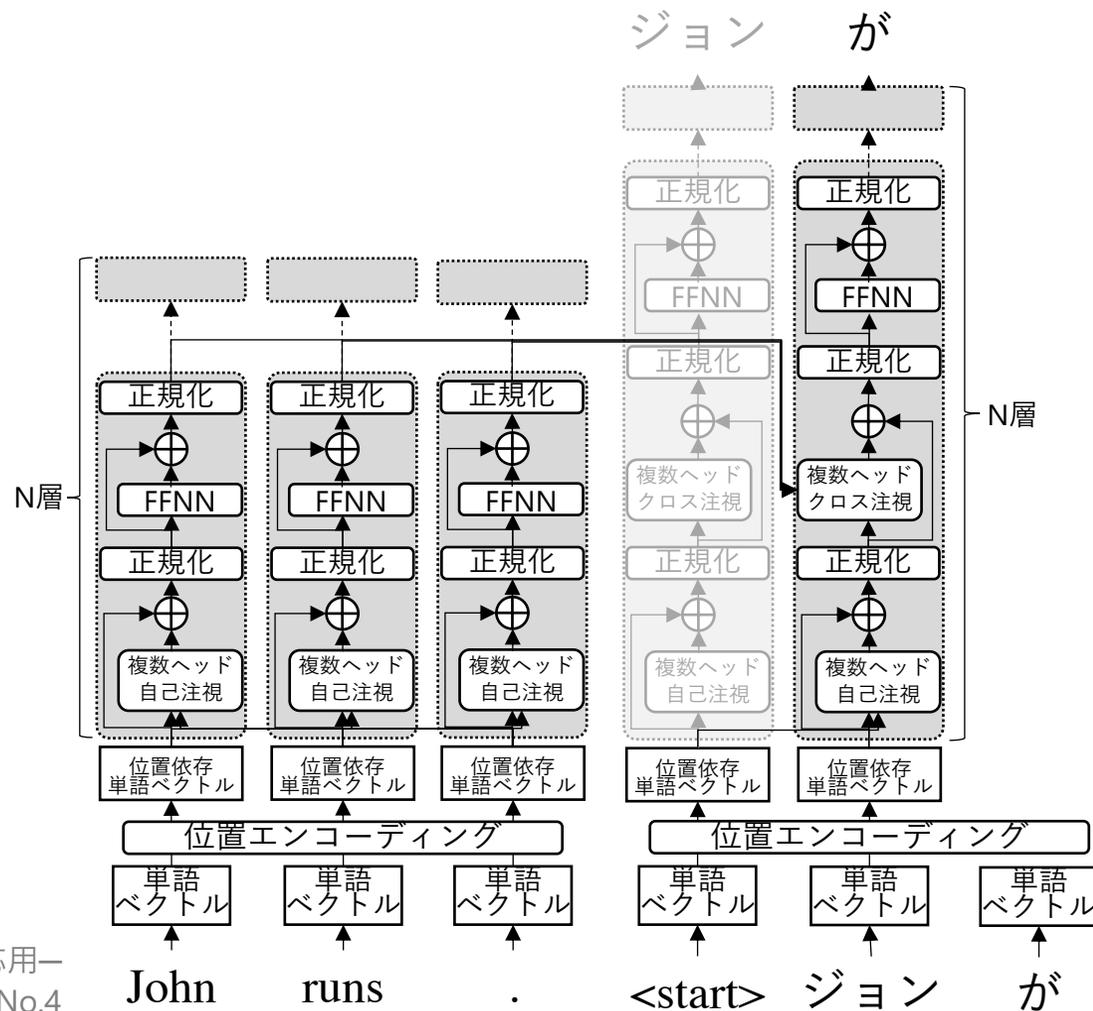


【ニューラル機械翻訳】

- 記号の対応が不明確
- 記号間の類似性が表現可能？

現在の標準モデル: Transformer (Vaswani+ 2017)

- 自己注視 (self-attention)
 - 入力相互の関係
 - 出力相互の関係
- 類似した計算を積み重ねる
 - 「層」…「深層学習」
- 一見複雑だが高速・高精度
 - 並列計算の恩恵



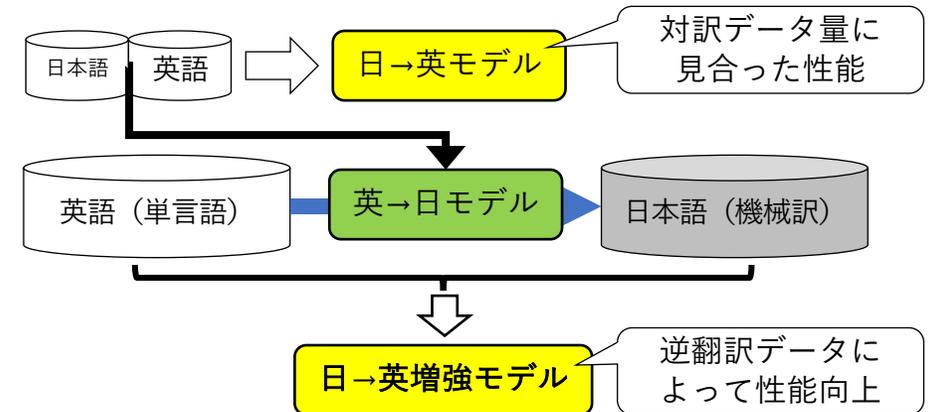
須藤, ニューラル機械翻訳の進展—系列変換モデルの進化とその応用—
人工知能学会誌 Vo.34, No.4

ニューラル機械翻訳の課題

- 訳抜け
 - 入力文の語句が訳されずに抜け落ちる
 - 特に学習が不十分な語句や難しい語句が抜けやすい
- 重複訳
 - 局所的な語句の訳出を繰り返す
 - 入力文にノイズが含まれると起こりやすい
- 内容の捏造
 - 入力文にないことが勝手に補われる

一般的に行われる工夫

- 単語より短い単位（サブワード）への分割
 - Sennrich+ (2016a), Kudo+ (2018) 等
 - internationalization → inter national ization
 - realization → real ization
- 逆翻訳…対訳コーパスの生成
 - Sennrich+ (2016b)等
 - 単言語コーパスの活用



特許翻訳の課題と 関連する機械翻訳技術の動向

特許文書の特徴／特許翻訳の課題

- 内容が専門的
- 文体がやや特殊
- 公的な文書
- 文の構造が複雑で長い
- 大量の翻訳処理が必要
- 言語によってはデータ不足（東南アジア言語）

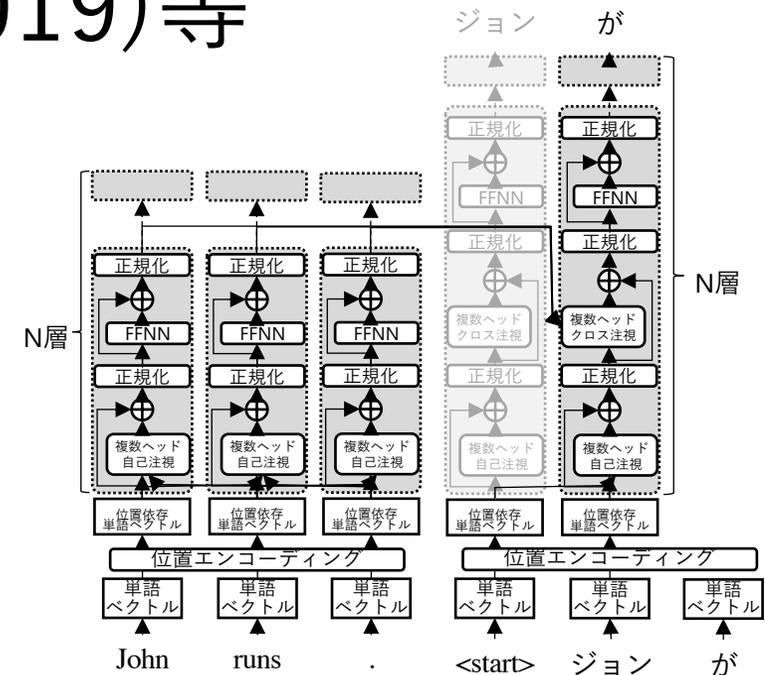
技術動向調査の中間報告

- デプロイ対策
 - 高速化や省メモリ化の取り組み
- 低リソース言語対策
 - 様々なデータや他言語の活用
- 自動評価（須藤も一部担当）
 - 深層学習に基づく訳質評価

※ 特許翻訳そのものを題材とした研究は現在少ない

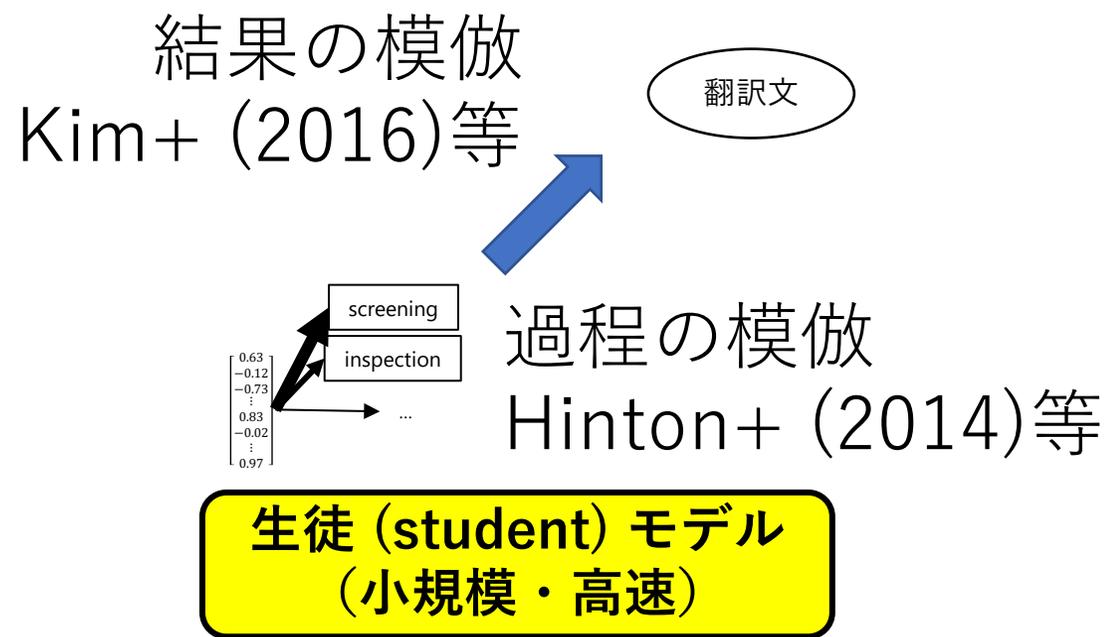
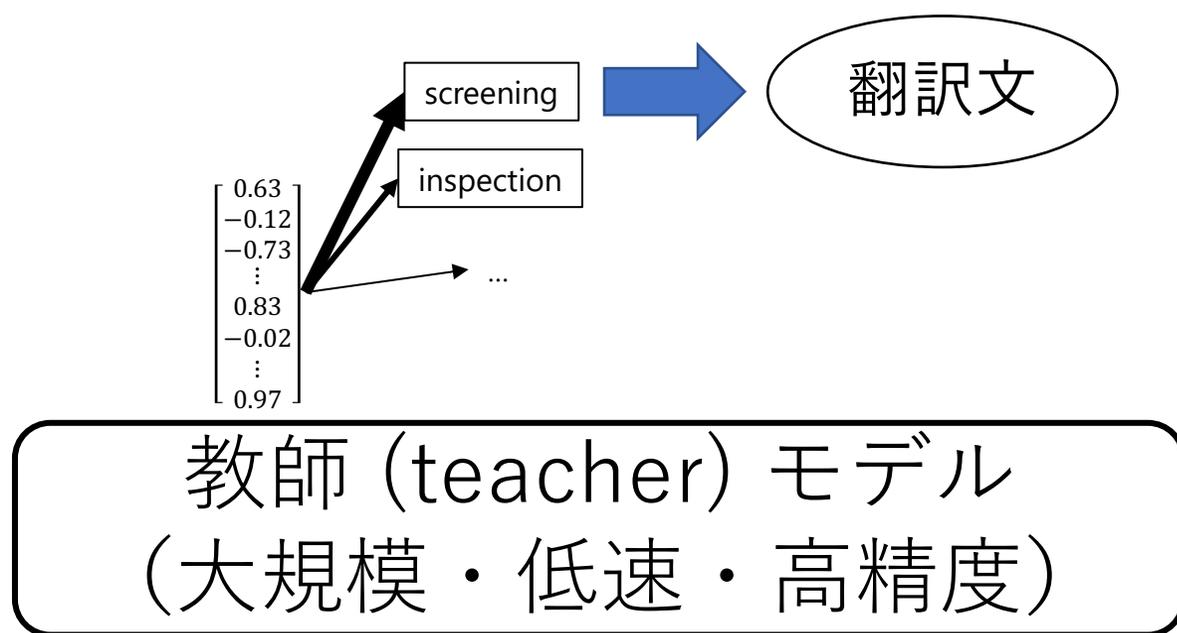
デプロイ対策①: モデルの軽量化

- ベクトル次元数／層数の削減 → 精度も低下
- 全く同じ計算層を繰り返し適用 → 省メモリ化
 - Dehghani+ (2018), Dabre+ (2019)等
 - 「パラメータ共有」



デプロイ対策①: モデルの軽量化 (続き)

- 「知識蒸留」(Knowledge Distillation)
- Teacher-Studentモデルとも
- 大きく低速なモデルを小さく高速なモデルで模倣



デプロイ対策②: ハードウェア特性の活用

- GPU・TPUによる並列化・高速化
- 浮動小数点の半精度(fp16)化
 - HW側の対応も進み単精度(fp32)より倍以上高速?
 - GoogleのTPUでは指数部8bitのbfloat16も
- 整数への量子化
 - Devlin+(2017), Bhandare+(2018)等
 - 行列の乗算部を整数(16bit/8bit)で量子化
 - CPU実行時に恩恵

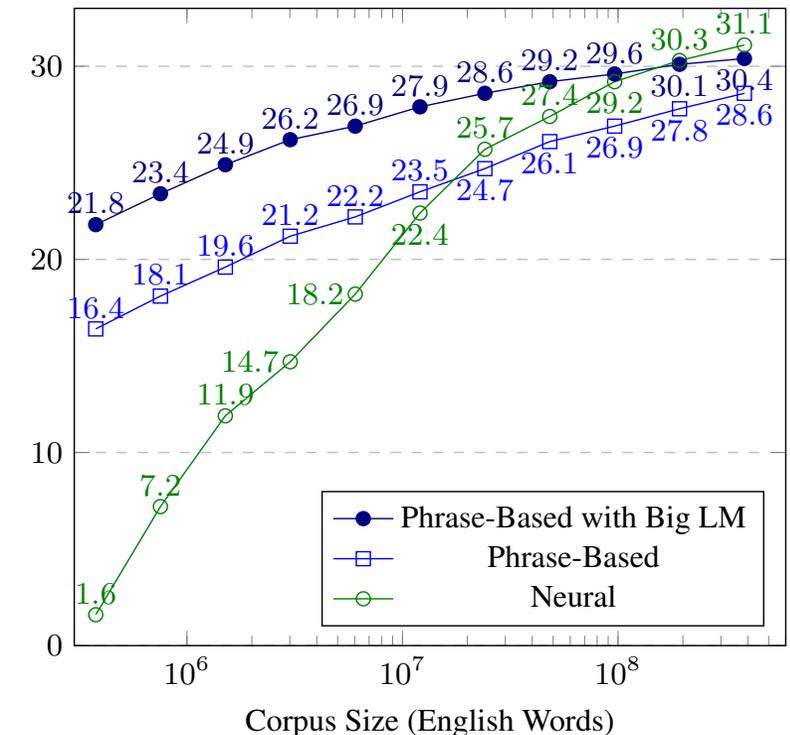
デプロイ対策③: 実装

- C/C++で実装
 - Python等より高速
 - PyTorch等の深層学習フレームワークは使えない
 - 緻密なメモリ管理が必要
 - Marian NMT <https://marian-nmt.github.io/>
- 途中計算の再利用 (キャッシュ)
 - Klein+ (2020), Hu+ (2020)等

低リソース言語対策①: NMTは不利か?

- ニューラル機械翻訳では低リソースは不利?
 - Koehn+ (2017)で提起
- 適切な設定により改善報告
 - Sennrich+ (2019), Duh+ (2020)
 - ベクトル正規化
 - パラメータ正則化
 - スムージング等

BLEU Scores with Varying Amounts of Training Data



低リソース言語対策②: ピボット翻訳

- 典型的には英語で仲介: 言語A→英語→言語B
 - 英語との対訳コーパスは比較的得やすい
 - 統計的機械翻訳の時代から知られたアプローチ
 - ニューラル機械翻訳では Costa-jussà+ (2018)等

低リソース言語対策③: 転移学習

- 高リソース言語で事前学習した後に低リソース言語で追加学習
 - 対象言語が類似していることが好ましい (語順等)
 - Nguyen+ (2017), Murthy+ (2019), Dabre+ (2019) 等

低リソース言語対策④: 多言語翻訳

- 複数の言語を翻訳できる単一の翻訳モデル
 - Many-to-Many翻訳: Firat+ (2016)
 - 様々な言語対でパラメータを共有したモデルを学習
 - Zero-shot翻訳: Johnson+ (2017)
 - 学習していない言語方向の翻訳が可能
 - 言語タグ (<2en>等) の使用
- マルチタスク学習の一種と見てよい

自動評価: 大きな分類

- モデルフリー手法
 - 表層のみを利用
 - 表層以外の情報も活用
- 汎用モデルベース手法
 - 追加学習が不要（ハイパーパラメータは除く）
- 専用モデルベース手法
 - 追加学習が必要

自動評価①: モデルフリー手法

- BLEU (Papineni+ 2002), NIST (Doddingon 2002)
 - 単語N-gram一致率に注目。NISTは頻度重みあり。
- TER (Snover+ 2006)
 - 編集距離に基づく。複数語のシフトを許容。
- IMPACT (Echizen'ya+ 2007)
 - 最長共通部分列を利用。
- RIBES (Isozaki+ 2010)
 - 語順の相関を利用。

自動評価①: モデルフリー手法 (続き)

- METEOR (Banerjee+ 2005)
 - 共通語幹や同義語を許容.
- MEANT (Lo 2011)
 - 述語項構造を利用 (半自動) .
- chrF (Popovic 2015), EED (Stanchev+ 2019)
 - 文字レベルで評価.

自動評価①: モデルフリー手法 (続き)

• 利点

- 学習が不要
- 高速
- 簡単に評価方法が共有可能
- 大部分は外部リソース不要 & 言語非依存
 - METEORは同義語辞書や語幹マッチルール, MEANTは解析器が必要

• 欠点

- 絶対的な精度は今ひとつ

自動評価②: 汎用モデル手法

- YiSi (Lo 2019)
 - 文中の意味役割と単語埋め込みを利用.
- WE_WPI (Echizen'ya+ 2019)
 - 単語埋め込みと Earth Mover's Distance を利用
- BERT_R (Mathur+ 2019), BERTScore (Zhang+ 2020)
 - BERTの文脈つきトークン埋め込みを利用
- MoverScore (Zhao+ 2020)
 - BERTの埋め込みと Earth Mover's Distanceを利用

自動評価②: 汎用モデル手法 (続き)

• 利点

- 厳密な表層マッピングが不要
 - METEORのような辞書的資源がなくてもよい
- 学習済みの単語埋め込み等の資源を活用可能
- 評価アノテーションつきデータでの追加学習が不要
 - ハイパーパラメータチューニング程度は必要なことも
- 言語依存性・ドメイン依存性が低い

• 欠点

- 絶対的な精度は専用モデルには及ばない

自動評価②: 専用モデル手法

- BEER (Stanojevic+ 2014)
 - 線形回帰モデルを利用.
- BLEND (Ma+ 2017)
 - 種々の自動評価手法のスコアに基づくサポートベクタ回帰(SVR)を利用

自動評価②: 専用モデル手法 (続き)

- RUSE (Shimanzuka+ 2018)
 - 文埋め込みに基づく回帰モデルを利用。
- BERT Regressor (嶋中+ 2019)
 - BERTの埋め込みに基づく回帰モデルを用い、ファインチューニングを行う。
- BLEURT (Sellam+ 2020)
 - BERTを言語推論などの補助タスクでファインチューニングし、さらに評価タスクでファインチューニング

自動評価②: 専用モデル手法 (続き)

- 利点

- 絶対的な精度の高さ

- 欠点

- 学習データ依存性が高い
 - ドメイン依存性
 - 品質依存性 (低訳質データで学習しても高訳質データの評価ができる保証がない)
 - BLEURTでは補助タスクの効果により依存性が低いと主張
- 学習時も評価時も計算が重くなりがち

自動評価③: 自動評価法の「メタ評価」

- ピアソンの積率相関係数の信頼性への疑問 (Mathur+ 2020)
 - 外れ値の影響を受けやすい
 - 性能が近接したシステム間の相関の問題
 - 小さな差は統計的有意であったとしても真に信頼できない

謝辞

- 調査報告内容の取りまとめについて、
AAMT/Japio特許翻訳研究会関係者に感謝します
 - 特に今回報告した各課題担当者の皆様
(デプロイ対策) 今村賢治さん, 園尾聡さん
(低リソース対策) 江原暉将さん, 今村賢治さん
(自動評価) 越前谷博さん

イベント告知

第6回特許情報シンポジウム（オンライン）

- 2021年2月26日（金）午後 開催
- 招待講演およびパネルディスカッションを予定
 - 特許・知財行政
 - 特許・知財実務
 - 自然言語処理・情報検索等研究開発
- 詳細は後日研究会Webサイトに掲載予定
 - <http://aamtjapio.com/>

文献リスト

ニューラル機械翻訳全般 (1/2)

- D. Bahdanau et al., [Neural Machine Translation by Jointly Learning to Align and Translate](#), Proc. ICLR (2015)
- K. Cho et al., [Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation](#), Proc. EMNLP (2014)
- S. Edunov et al., [Understanding Back-Translation at Scale](#), Proc. EMNLP (2018)
- T. Kudo, [Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates](#), Proc. ACL (2018)
- M.-T. Luong et al., [Effective Approaches to Attention-based Neural Machine Translation](#), Proc. EMNLP (2015)

ニューラル機械翻訳全般 (2/2)

- R. Sennrich et al., [Improving Neural Machine Translation Models with Monolingual Data](#), Proc. ACL (2016a)
- R. Sennrich et al., [Neural Machine Translation of Rare Words with Subword Units](#), Proc. ACL (2016b)
- I. Sutskever et al., [Sequence to Sequence Learning with Neural Networks](#), Proc. NIPS (2014)
- A. Vaswani et al., [Attention Is All You Need](#), Proc. NIPS (2017)

デプロイ対策

- A. Bhandare et al., [Efficient 8-Bit Quantization of Transformer Neural Machine Language Translation Model](#), Proc. ICML Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (2019)
- R. Dabre et al., [Recurrent Stacking of Layers for Compact Neural Machine Translation Models](#), Proc. AAAI (2019)
- M. Dehghani et al., [Universal Transformers](#), Proc. ICLR (2019)
- J. Devlin, [Sharp Models on Dull Hardware: Fast and Accurate Neural Machine Translation Decoding on the CPU](#), Proc. EMNLP (2017)
- G. Hinton et al., [Distilling the Knowledge in a Neural Network](#), Proc. NIPS 2014 Deep Learning Workshop (2014)
- Y. Kim et al., [Sequence-level Knowledge Distillation](#), Proc. EMNLP (2016)

低リソース言語対策 (1/2)

- M. Costa-jussa et al., [English-Catalan Neural Machine Translation in the Biomedical Domain through the cascade approach](#), Proc. LREC Workshop MultilingualBIO: Multilingual Biomedical Text Processing (2018)
- R. Dabre et al., [Exploiting Multilingualism through Multistage Fine-Tuning for Low-Resource Neural Machine Translation](#), Proc. EMNLP-IJCNLP (2019)
- K. Duh et al., [Benchmarking Neural and Statistical Machine Translation on Low-Resource African Languages Cross-over](#), Proc. LREC (2020)
- O. Firat et al., [Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism](#), Proc. NAACL (2016)
- M. Johnson et al., [Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation](#), Transactions of ACL (2017)

低リソース言語対策 (2/2)

- P. Koehn et al., [Six Challenges for Neural Machine Translation](#), Proc. of WNMNT (2017)
- R. Murthy et al., [Addressing word-order Divergence in Multilingual Neural Machine Translation for extremely Low Resource Languages](#), Proc. NAACL-HLT (2019)
- T. Q. Nguyen et al., [Transfer Learning across Low-Resource, Related Languages for Neural Machine Translation](#), Proc. IJCNLP (2017)
- R. Sennrich et al., [Revisiting Low-Resource Neural Machine Translation: A Case Study](#), Proc. ACL (2019)

自動評価 (1/3)

- S. Banerjee et al., [METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments](#), Proc. WS on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization (2005)
- C.-K. Lo et al., MEANT: [An inexpensive, high-accuracy, semi-automatic metric for evaluating translation utility based on semantic roles](#), Proc. ACL (2011)
- M. Popović, [CHRf: character n-gram F-score for automatic MT evaluation](#), Proc. WMT (2015)
- P. Stanchev et al., [EED: Extended Edit Distance Measure for Machine Translation](#), Proc. WMT (2019)
- C.-K. Lo, [YiSi - a Unified Semantic MT Quality Evaluation and Estimation Metric for Languages with Different Levels of Available Resources](#), Proc. WMT (2019)

自動評価 (2/3)

- H. Echizen'ya et al., [Word Embedding-Based Automatic MT Evaluation Metric using Word Position Information](#), Proc. NAACL-HLT (2019)
- N. Mathur et al., [Putting Evaluation in Context: Contextual Embeddings Improve Machine Translation Evaluation](#), Proc. ACL (2019)
- T. Zhang et al., [BERTScore: Evaluating Text Generation with BERT](#), Proc. ICLR (2020)
- W. Zhao et al., [MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance](#), Proc. EMNLP-IJCNLP (2019)
- M. Stanojević et al., [BEER: BEtter Evaluation as Ranking](#), Proc. WMT (2014)
- Q. Ma et al., [Blend: a Novel Combined MT Metric Based on Direct Assessment — CASICT-DCU submission to WMT17 Metrics Task](#), Proc. WMT (2017)

自動評価 (3/3)

- H. Shimanaka et al., [RUSE: Regressor Using Sentence Embeddings for Automatic Machine Translation Evaluation](#), Proc. WMT (2018)
- H. Shimanaka et al., [Machine Translation Evaluation with BERT Regressor](#), arXiv preprint (2019)
- T. Sellam et al., [BLEURT: Learning Robust Metrics for Text Generation](#), Proc. ACL (2020)
- N. Mathur et al., [Tangled up in BLEU: Reevaluating the Evaluation of Automatic Machine Translation Evaluation Metrics](#), Proc. ACL (2020)