

# IJCNLP2013 参加報告 (その2)

## – Best Paper と機械翻訳の論文 –

中澤 敏明<sup>†</sup>

### 1 はじめに

本稿では IJCNLP2013 の Best Paper に選ばれた論文の概要と、機械翻訳に関する論文を数本、簡単に紹介する。

### 2 SuMT: A Framework of Summarization and MT

IJCNLP2013 の Best Paper Award は言語横断文書要約 (cross-language summarization) についての研究である “SuMT: A Framework of Summarization and MT” (Bouamor, Mohit, and Oflazer 2013) に贈られた<sup>1</sup>。Carnegie Mellon University in Qatar のグループによる研究で、英語の文書をアラビア語で高精度に要約する研究である。システムの概要を図 1 に示す。全体として以下のような手順で要約を生成する。

- (1) 入力の英語文書を文ごとに機械翻訳でアラビア語に翻訳 (図 1 の (a))
- (2) 翻訳品質を自動推定し、スコア付け (図 1 の (b))
- (3) 翻訳品質を加味しつつ**入力の英語文書**を extractive に要約し、これに対応するアラビア語翻訳を出力することで、アラビア語による要約を生成 (図 1 の (c))

このうちポイントは 2 と 3 であり、各文の情報量と翻訳品質の両方を考慮することで、より良い要約を生成することに成功している。以下、(b) と (c) の部分についてさらに詳しく説明する。

<sup>†</sup> 科学技術振興機構, Japan Science and Technology Agency

<sup>1</sup> ちなみに、SuMT は [sám̥ti:] と発音するらしい

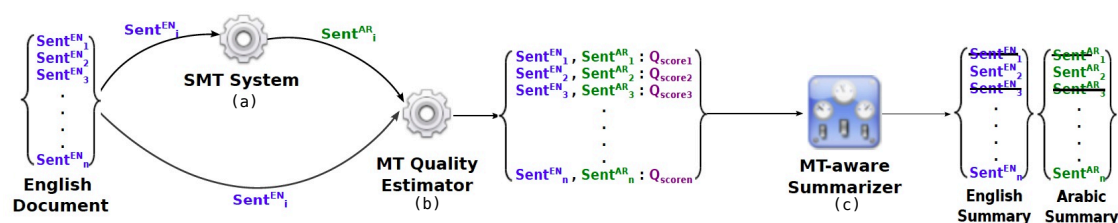


図1 SuMTの概要

## 2.1 機械翻訳品質の自動推定

機械翻訳の品質推定はそれだけで一つの研究分野であり、機械翻訳ワークショップのWMT<sup>2</sup>でもタスクの一つとなっているぐらいである。機械翻訳はただでさえ評価が難しく、品質推定となるとさらに難しい課題であり、これまでも様々な方法が提案されている<sup>3</sup>。

さて、この論文ではSVMを用いた手法(Mohit and Hwa 2007)により品質推定を行っている。SVMのトレーニングには入力文とその機械翻訳文のペアに対し、翻訳の質が良いか悪いかの二値の正解ラベルが付与されたデータを用いる。正解ラベルを人手で付与するのは大変なので、ここでは参照訳と機械翻訳の自動評価指標である Translation Edit Rate(TER)(Snover, Dorr, Schwartz, Micciulla, and Makhoul 2006)<sup>4</sup>を用いて、以下のように自動で正解ラベルを付与している:

- (1) 翻訳された文書全体のTERを測る
- (2) 各文ごとのTERを測り、文TERが文書TERより低ければ良い翻訳、高ければ悪い翻訳というラベルを付与する

つまり、文書全体の翻訳の中には良い翻訳と悪い翻訳が混在しており、全体の翻訳精度はおおよそ各文の翻訳精度の平均となっているため、文の翻訳精度が全体の精度より高い場合は良い翻訳、そうでなければ悪い翻訳と仮定しているのである。

SVMの素性として、以下のようなものを用いている。

- General features: 単語数、入力と出力の単語数の比、句読点、数字、内容語
- Language model scores: 出力文の言語モデルによる尤度
- MT-based scores: 出力文中の未知語数と割合、出力文でdetokenizationが行われた割合
- Morphosyntactic features: 入力文と出力文でのPOSタグ順の違い、名詞・動詞・固有名詞の割合

<sup>2</sup><http://www.statmt.org/wmt14/>

<sup>3</sup>実は機械翻訳ユーザの立場からは翻訳の確信度のようなものを出力してほしいという要望はかなり強いので、品質推定が実現できれば社会へのインパクトはかなり大きいと思われる。

<sup>4</sup>TERは参照訳と機械翻訳の出力との類似度を編集距離に基づいて評価するのだが、通常の編集距離が「挿入」「削除」「置換」の3つの操作を用いて測るのに対し、TERではさらに「シフト」という操作を加え、並び替えに対する制約を弱めている。

- Document-level features: 上記の素性を文書レベルで平均したもの

## 2.2 翻訳の質を考慮した要約

文書要約の手法は大きく分けて abstractive な手法と extractive な手法がある。abstractive な手法では、入力文書から新たな文を生成しながら要約を行い、extractive な手法では入力文書の文から重要な文を選択することで要約を行う。この論文では文書要約のベースラインシステムとして extractive な手法で要約を行う MEAD(Radev, Allison, Blair-Goldensohn, Blitzer, Celebi, Dimitrov, Drabek, Hakim, Lam, Liu, Otterbacher, Qi, Saggion, Teufel, Topper, Winkel, and Zhang 2004) を利用している。MEAD は文の長さ、出現位置、重要な単語を含む割合などを素性として利用する。SuMT ではこれにさらに翻訳の品質を考慮して要約を行う。各素性に重みを掛けて線形和を取ることで各文のスコアを計算し、スコアの高い文から順に選択することで要約を生成する。MEAD でも使われている基本素性の重みはデフォルトの値を用いるが、翻訳品質素性の重みは、開発データを用いて BLEU を目的関数として最適化している。

## 2.3 実験と結果

実験には機械翻訳ワークショップ NIST で使われている英語-アラビア語対訳コーパスを利用している。またこのコーパスのテストデータの英語側とアラビア語側から独立に、各言語の母語話者に要約を作ってもらい、要約の正解データとして利用している。機械翻訳システムとしてフレーズベース SMT のツールキット Moses を利用し、品質推定のための SVM として LIBSVM を利用している。

評価は機械翻訳としての評価と、要約としての評価の2つを行っている。機械翻訳評価では、様々な要約システムにより英語文書を要約し、それに対応するアラビア語機械翻訳結果と参照訳を用いて、BLEU・TER・METEOR で評価している。どの指標においても提案手法である SuMT が最もよい精度を達成しているが、注意すべき点として要約システムごとに要約として選択する文が異なるため、必然的に機械翻訳評価に用いられる文セットが異なる。

要約としての評価は、英語側とアラビア語側を独立に、要約の精度評価で一般的に用いられている ROUGE で評価している。英語側の要約精度はベースラインの MEAD が最も高かったが、アラビア語側の精度は提案手法である SuMT が最も高い結果となり、翻訳品質を考慮することでより良い要約が生成できることが示された。

## 3 機械翻訳関係の論文の紹介

ここでは個人的に興味を持った機械翻訳に関する論文を、いくつか簡単に紹介する。

Meta-level Statistical Machine Translation(Ebrahimi, Meshgi, Khadivi, and Shiri Ahmad Abady

2013) は Stacked Generalization または Stacking と呼ばれる機械学習の枠組みを機械翻訳に適用したという論文である。トレーニング対訳コーパスを  $N$  分割交差検定することで、 $N$  個の異なる翻訳システム (base-level SMT) とそれぞれの翻訳システムによる残りのデータの翻訳を生成する。この翻訳結果と元の対訳コーパスの目的言語側とを組み合わせると新たな対訳コーパスとし (このコーパスは原言語も目的言語も同じ言語となる)、これを用いて別の翻訳システム (meta-level) を構築する。新たな文を翻訳する際には、元のトレーニング対訳コーパス全体を用いて構築した翻訳システムで一旦目的言語に翻訳し、これをさらに meta-level システムで翻訳 (後修正のようなもの) することで、最終的な出力を得る。

Hypothesis Refinement Using Agreement Constraints in Machine Translation (Gandhe and Gangadharaiyah 2013) は言語モデルやフレーズペアだけでは扱いきれないような、遠い位置にある語に影響される動詞の活用などを正しく翻訳するという論文である。入力文の情報から、要求される出力文の主語・目的語・動詞の形を推定し、これに合致しないフレーズペアを翻訳時に削除したりペナルティーをかけるなどする。また要求される形がどのフレーズペアにも存在しない場合には、新たに適切な活用を含むフレーズペアを生成して利用する。

Ensemble Triangulation for Statistical Machine Translation (Razmara and Sarkar 2013) は原言語から目的言語の翻訳において、直接の翻訳と中間言語を介した翻訳を組み合わせるという論文だが、中間言語として複数の言語を利用する方法を提案している。またフレーズテーブルを組み合わせる方法もいくつか検討している。

Selective Combination of Pivot and Direct Statistical Machine Translation Models (El Kholy, Habash, Leusch, Matusov, and Sawaf 2013) も中間言語を利用した翻訳についての論文だが、こちらは巨大になりがちな中間言語を介した場合のフレーズテーブルを、精度を落とさずに小さくするための方法を検討している。

## 4 まとめ

機械翻訳に関する論文は本稿で紹介したもの以外にも、翻訳の素性選択に関するものや言語モデルに関するものなど、様々な研究発表があり非常に面白かった。我々のグループからもトピックモデルを用いたコンパラブルコーパスからの翻字マイニングに関する研究 (Richardson, Nakazawa, and Kurohashi 2013) とコンパラブルコーパスから対訳断片を高精度に獲得する研究 (Chu, Nakazawa, and Kurohashi 2013) の発表を行ったので、興味のある方はぜひ目を通していただけたい。

## 参考文献

- Bouamor, H., Mohit, B., and Oflazer, K. (2013). “SuMT: A Framework of Summarization and MT.” In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 270–278 Nagoya, Japan. Asian Federation of Natural Language Processing.
- Chu, C., Nakazawa, T., and Kurohashi, S. (2013). “Accurate Parallel Fragment Extraction from Quasi-Comparable Corpora using Alignment Model and Translation Lexicon.” In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 1144–1150 Nagoya, Japan. Asian Federation of Natural Language Processing.
- Ebrahimi, S., Meshgi, K., Khadivi, S., and Shiri Ahmad Abady, M. E. (2013). “Meta-level Statistical Machine Translation.” In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 1151–1157 Nagoya, Japan. Asian Federation of Natural Language Processing.
- El Kholy, A., Habash, N., Leusch, G., Matusov, E., and Sawaf, H. (2013). “Selective Combination of Pivot and Direct Statistical Machine Translation Models.” In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 1174–1180 Nagoya, Japan. Asian Federation of Natural Language Processing.
- Gandhe, A. and Gangadharaiyah, R. (2013). “Hypothesis Refinement Using Agreement Constraints in Machine Translation.” In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 429–437 Nagoya, Japan. Asian Federation of Natural Language Processing.
- Mohit, B. and Hwa, R. (2007). “Localization of Difficult-to-Translate Phrases.” In *Proceedings of the Second Workshop on Statistical Machine Translation*, pp. 248–255 Prague, Czech Republic. Association for Computational Linguistics.
- Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Celebi, A., Dimitrov, S., Drabek, E., Hakim, A., Lam, W., Liu, D., Otterbacher, J., Qi, H., Saggion, H., Teufel, S., Topper, M., Winkel, A., and Zhang, Z. (2004). “MEAD — A platform for multidocument multilingual text summarization.” In *Conference on Language Resources and Evaluation (LREC)* Lisbon, Portugal.
- Razmara, M. and Sarkar, A. (2013). “Ensemble Triangulation for Statistical Machine Translation.” In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 252–260 Nagoya, Japan. Asian Federation of Natural Language Processing.
- Richardson, J., Nakazawa, T., and Kurohashi, S. (2013). “Robust Transliteration Mining from Comparable Corpora with Bilingual Topic Models.” In *Proceedings of the Sixth Interna-*

*tional Joint Conference on Natural Language Processing*, pp. 261–269 Nagoya, Japan. Asian Federation of Natural Language Processing.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). “A study of translation edit rate with targeted human annotation.” In *In Proceedings of Association for Machine Translation in the Americas*, pp. 223–231.

## 略歴

**中澤 敏明（正会員）：**

2010年京都大学大学院情報学研究科知能情報学専攻博士後期課程修了。博士（情報学）。機械翻訳の研究に従事。

(2013年11月13日依頼)

(2014年1月22日受付)