EMNLP2013 参加報告（その３）
– 統計的機械翻訳関係で面白そうな論文をいくつか紹介 –

乗松潤矢[†]

1　はじめに

統計的機械翻訳において、性能の鍵を握るのはデータ量とモデルである。データ量が少ないとすぐれたモデルであっても過学習の危険が高くなる。また、データ量が多くてもモデルが不適切では性能は期待できない。

本報告では、EMNLP 2013 で発表された論文のうち、機械翻訳に関する興味深い論文を 2 本紹介する。一方は、対訳コーパスの文数が限られている場合での性能向上を目指した論文であり、他方は、IBM モデルのグローバル最適化による単語アライメントの改善を目指した論文である。

本報告を通して、これらの論文に興味を持っていただけるなら幸いである。

2　少ない対訳コーパス下での統計的機械翻訳

統計的機械翻訳では、そこそこの性能を出すために大規模な学習データが必要となる。しかし単言語コーパスならまだしも、大規模な対訳コーパスを集めることは現実的になかなか難しく、筆者を含む統計的機械翻訳研究者・開発者にとって一番の悩みの種になっている。

過去にも言語モデルの学習データ量を増加させると翻訳性能が向上することが示されたことがあった (Brants, Popat, Xu, Och, and Dean 2007) が、そちらの方は目的言語側の話であり、今回は翻訳モデルの研究として原言語側も使うという意味で一線を画している。

EMNLP2013 において、単言語コーパスをうまく使うことで対訳コーパスを補完させる研究が発表されていた。Qing Dou and Kevin Knight “Dependency-Based Decipherment for Resource-Limited Machine Translation” がそれである (Dou and Knight 2013)。この手法により、数%の BLEU 値向上を達成できたとのことで、対訳コーパスにしばしば悩まされる筆者としては大変興味深く、かつきわめて有用と思われる。

[†]筑波大学, The University of Tsukuba

Dou らは以前にも同じコンセプトの論文 (Dou and Knight 2012) を発表しており本論文はその改良である。過去の論文では、未知の原言語 bigram に対して、目的言語 bigram を “decipher” するというものだった。“decipher” というからには未知の原言語は目的言語から暗号化されたものであるとして復号を試みる。すなわち原言語 bigram f_1, f_2 は、以下のように確率的に生成されたとする。

$$P(f_1, f_2) = \sum_{e_1 e_2} P(e_1, e_2) P(f_1|e_1) P(f_2|e_2)$$

ここで、e_1, e_2 は、あらゆる目的言語単語であり、$P(e_1, e_2)$ は目的言語の言語モデルによって与えられる。この式を使えば、原言語 bigram の集合に対して尤度を計算できるため、$P(f|e)$ が推定できる[1]。

勘のいい方はすでにお気づきだろうが、Dou らの過去の研究では、単語のリオーダリングが考慮されていない。これでは英日翻訳のような語順変化のある言語対には対応できない。そこで、本論文ではかつての bigram の変わりに、依存構造 bigram を利用することで単語リオーダリングに対応した。依存構造 bigram とは、すなわち係り受けによる 2 単語連鎖である。依存構造解析を行うと、離れた単語の関係性が明らかになるが、ここで係り元と係り先の関係で bigram を作るのである。依存構造 bigram であれば語順が入れ変わっても両言語間で対応が取れるはずである。

実験結果を見ると、この改善により、単語対応の精度が 500%(!) も向上しているのがわかる。ただ、500%向上とはいえ、実際の正答率は 30%ない程度で、まだまだ研究の余地はありそうな印象を受けるが、翻訳実験では、BLEU が最大で 1.8%も向上しており、確かに翻訳精度にも影響を与えている。翻訳の際には、

- 対訳コーパスに混ぜる
- 対訳コーパスとは別に持っておき、未知語に対して翻訳候補として使う
- 翻訳モデルとは別の対訳辞書として保持しておき、デコーディング時には翻訳モデルと同時に参照する

の 3 通りを行いこの順に翻訳性能が向上しているのも興味深い。

本研究は統計的機械翻訳の対訳コーパス問題を解決する糸口として、今後の研究の進展に期待している。

[1] 普通に EM アルゴリズムを使うと、計算量が膨大になるらしく、Dou らは gibbs sampling により推定を行っている。

3 単語アライメントのグローバル最適化

統計的機械翻訳において、数多くの翻訳モデルが提案されているが、その中でも翻訳モデルの祖とされるのが IBM モデル (Brown, Pietra, Pietra, and Mercer 1993) である。句を単位とする翻訳が主流となった現代でも、IBM モデルはなお単語アライメントの基礎理論として統計的機械翻訳を支えている (Och and Ney 2004)。

次に紹介するのは Andrei Simion, Michael Collins, Clifford Stein "A Convex Alternative to IBM Model 2" である (Simion, Collins, and Stein 2013)。タイトルからわかる通り、本論文で取り扱うのは IBM モデル 2 である。IBM モデルは 1 から 5 まで提案されているが、数字が大きくなるにしたがって複雑さも増していく。IBM モデル 1 はその中でも最もシンプルなモデルであり、数理的にも美しいモデルであるが、単語の並べ替えをほとんど考慮していないため性能は低い。IBM モデル 2 では単語の並べ替えを考慮しているため、式は複雑になるもののモデル 1 に比べると性能が高い。著者らによると、IBM モデル 2 の問題点は、関数が凸関数でないことにあるという。一般に、目的関数が凸関数でない場合は、EM アルゴリズムではグローバルの最適解を保証できない。IBM モデル 2 でも EM アルゴリズムが用いられるが、この問題があるために従来の推定結果が最適解である保証はない。そこで、著者らは IBM モデル 2 に対して 2 つの修正を加えることで、より性質の良いモデルを提案している。

一つ目の修正は、目的関数の緩和である。本論文は、この IBM モデル 2 の目的関数を凸関数となるように緩和する。この緩和問題は凸関数であるがゆえにグローバルでの最適解を与えることができ、結果としてモデル性能の向上につながる。

IBM モデル 2 では、単語翻訳確率 t と歪み確率 d を考えるため、目的関数に $t \times d$ の項が表れる。著者らによると、この項により、目的関数が凸でなくなっている。そこで彼らは別の変数 q を用意し、

$$\begin{cases} q < t \\ q < d \end{cases}$$

という条件を追加する。さらに、目的関数の $t \times d$ となっていた部分を q に置き換えることで、目的関数を凸関数に緩和する。最適化問題としては、t, d, q の 3 種類の変数を最適にする値を求める[2]。これが基本アイディアである。

2 つ目は、IBM モデル 1 の利用である。著者らが問題視したのは、この緩和法では、t と d のどちらが重要かを表現できず、d の方を中心に最適化が行われてしまう可能性がある点である。IBM モデル 2 では、歪み確率 d は文中の単語位置を用いてモデル化されるが、これは絶対位置を用いているため元々高い性能は期待できない。それよりも単語対応確率 t に対して高い精度

[2] ここでは説明のために簡略化して 3 変数で記載したが、実際には大量の変数が必要である。

を求める方が合理的である。そこで、t に重点を置くことを明示するために、IBM モデル 1 の目的関数を今回の目的関数に足し算する。歪み確率 d が含まれていない IBM モデル 1 の目的関数を足すことで、相対的に単語翻訳確率 t の重要性が増す。IBM モデル 1 の目的関数は凸関数であることが知られているため、この修正によって目的関数が凸であることは失われない。

上記二つの修正により、従来の EM アルゴリズムで求めたモデルパラメータよりも高い性能を示すという。実験においても、単語対応の精度が従来の IBM モデル 2 を上回る性能を示しており、本手法の有用性が見て取れる。一点残念な点を挙げるとすると、本論文では翻訳実験が行われていなかった点である。現代の統計的機械翻訳において、IBM モデル 2 の推定結果が全体の翻訳性能にどの程度影響を与えるのか、興味深い。もしかすると、翻訳システム全体としての性能向上のためには、IBM モデル 3,4 と同様の問題を解決する必要があるのかもしれない。そういった意味でも今後の研究に注目したい。

4　おわりに

本報告では、EMNLP 2013 で発表された論文の中から機械翻訳関連で筆者が興味深いと思った論文を紹介した。限られた言語資源のなかでどう性能を出していくかといった問題や、理論的には美しいが解が求まらない、といった問題は、機械翻訳に限らずその他の分野でもしばしば経験することだろう。今回紹介した論文は、それらに正面から挑んだもので、大変刺激を受けた。

いずれも筆者自身の研究とはほとんど関係なさそうな分野であり、会議に出席しなければ興味を持つことはなかった。引きこもって研究することも大切だが、こういった機会に外に目を向け、新しい世界を覗き見るのも大変楽しいと思わされた。

末筆ではあるが、本報告で紹介した内容はわかりにくい点も多かったと思うが、これは筆者の力不足によるところである。これらの論文に興味を持っていただける方がいらっしゃればぜひ各論文を参照し、その素晴らしさを味わっていただきたいと思う。

参考文献

Brants, T., Popat, A. C., Xu, P., Och, F. J., and Dean, J. (2007). "Large Language Models in Machine Translation." In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 858–867 Prague. Association for Computational Linguistics.

Brown, F. B., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). "The mathematics of statistical machine translation: Parameter estimation." *Computational linguistics*, **10598**.

Dou, Q. and Knight, K. (2012). "Large Scale Decipherment for Out-of-Domain Machine Translation." In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 266–275 Jeju Island, Korea. Association for Computational Linguistics.

Dou, Q. and Knight, K. (2013). "Dependency-Based Decipherment for Resource-Limited Machine Translation." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1668–1676 Seattle, Washington, USA. Association for Computational Linguistics.

Och, F. J. and Ney, H. (2004). "The alignment template approach to statistical machine translation." *Computational linguistics*, **30** (October 2003).

Simion, A., Collins, M., and Stein, C. (2013). "A Convex Alternative to {IBM} Model 2." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1574–1583 Seattle, Washington, USA. Association for Computational Linguistics.

略歴

乗松潤矢（学生会員）：2007 年筑波大学第三学群情報学類卒業、2009 年同大学院システム情報工学科コンピュータサイエンス専攻博士前期課程修了し、同年、株式会社平和情報センター (現、富士通エフ・アイ・ピー・システムズ株式会社) に入社。2012 年、同社を退職、筑波大学システム情報工学研究科コンピュータサイエンス専攻博士後期課程に入学。現在、フリーランスで生計をたてつつ研究活動に励む。

(2013 年 11 月 13 日依頼)
(2014 年 1 月 21 日受付)