

より現実的な自動要約に向けて

西川 仁

NTT メディアインテリジェンス研究所

nishikawa.hitoshi@lab.ntt.co.jp

1 はじめに

著者は自動要約の研究に携わってきた。2013年9月に情報処理学会第213回自然言語処理研究会にて自動要約に関する近年の動向および今後の課題について講演する機会があった[9]。講演では、現在の自動要約における課題は、外的な評価、大規模データの利用、および読み手の考慮の3点であることを述べた。おそらくこれらの課題に通底する方向性は、より現実的な状況を考えた自動要約システムというものになるだろう。本稿では、民間企業に勤務する研究者としての立場から、これらの課題にどのように取り組んでいるのか、また取り組んでいくべきであるか考えているのか述べてい

2 外的な評価

ROUGE [2] の発明以来、自動要約に関連する技術は長足の進歩を遂げた。内的な評価¹を用いれば短時間で手法の性能を評価することができ、結果としてより短い時間で手法の改善が可能になる。これは機械翻訳において BLEU [6] の発明がもたらした効果と同様である

一方、つねづね指摘されることではあるが、こういった内的な評価尺度の妥当性は要約の対象とするデータ毎に適切に吟味されなければならない[3]。また、内的な評価尺度に頼った要約評価を行うことで、本来の要約の目的、すなわち読み手が自動要約によって作成された要約を読むことで、原文書に含まれる重要な情報を短時間に理解することができたか、という目的が忘れ去られる恐れもある。

1つの解決策は、小規模でも、実際に外的な評価を行うことであろう。特に、要約のみを被験者に提示した上で、原文書の重要な事実を理解できたか問うといった評価を行うことによって、要約システムの開発者は

開発中のシステムに関する重要な示唆を得ることができよう。また、人手に依存せずに実際的な評価を行う方法もある。例えば野本らは情報検索課題を通じて要約の性能を評価をしている[5]。要約前の文書と要約後の文書を比較し、情報検索課題において精度の低下がみられなければ、適切な要約ができているという仮定は、要約の実際的な評価にふさわしいものである。

3 大規模データ

自動要約において用いられるコーパスの規模は必ずしも大きくない。Document Understanding Conference² や Text Analysis Conference³ , Text Summarization Challenge⁴ などで用いられている入出力対は数十から数百の規模であり、他の自然言語処理課題において用いられているラベル付きデータに含まれる入出力対の数と比べると少ない。比較的大規模なコーパスは Ziff-Davis コーパスであり、Marcu は Ziff-Davis コーパスに含まれる原文書と要約の 6,942 対を用いて文アライメントを行っている[4]。しかし、統計的機械翻訳において用いられる入出力対に比べるとまだまだ少ない。また、原文書、要約ともに全て英語であり、日本語での要約研究にはそのまま利用することができない。大規模なデータが存在しなければ、信頼のおける定量的評価が難しくなり、要約システムの正しい性能を評価することが難しくなる。

これに関する解決策は2つあるものと思われる。1つは、入出力ペアを自動的に獲得する方法を考案することである。Filippova は文短縮における訓練事例の不足を補う方法を提案している[1]。自動的なコーパスの構築によって、費用をかけることなく、大規模なコーパスを得ることができよう。もう1つは、実際に、人間の手によって作成されている何らかの要約

¹内的な評価、外的な評価の別については奥村らによる文献[8]を参照のこと。

²<http://duc.nist.gov/>

³<http://www.nist.gov/tac/>

⁴<http://lr-www.pi.titech.ac.jp/tsc/>

を入手し、利用することである。インターネット上には、人手によって作成された、ある種の要約といつてよいテキストが散見される。そういったテキストを収集し、要約研究に利用することで、大規模なデータが要約システムに与える影響を分析することができるだろう。

4 読み手の考慮

最後の問題は、読み手に関する問題である。現在の要約システムにおいては、典型的には、読み手から提示される、要約システムに対する要求は単語の集合の形で表現されることが多い。実際には読み手の要約器に関する要求はより複雑なものであると想定するのは自然であり、また読み手そのものを考慮することも必要だろう。例えば、読み手の性別や年齢、職業などは要約システムがよりよい要約を出力するための手がかりになりうる。例えば、「やさしい日本語」や関連する研究 [7, 10] が企図するように、テキストを読み手に合わせて変換することは、自動要約を含む、ユーザの情報アクセスを助ける技術にとって重要である。

5 終わりに

本稿では特に自動要約を例として述べたが、これらの課題は自動要約分野に限らず自然言語処理の広い領域においても課題であるようにも思われる。本稿の問題提起が今後の自然言語処理研究の方向性を考える上での一助となれば幸いである。

参考文献

- [1] Katja Filippova. Overcoming the lack of parallel data in sentence compression. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1481–1491, 2013.
- [2] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In *Proceedings of ACL Workshop Text Summarization Branches Out*, pp. 74–81, 2004.
- [3] Feifan Liu and Liu Yang. Correlation between rouge and human evaluation of extractive meeting summaries. In *Proceedings of the 46th Annual Conference of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), Short Papers*, pp. 201–204, 2008.
- [4] Daniel Marcu. The automatic construction of large-scale corpora for summarization research. In *Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 137–144, 1999.
- [5] Tadashi Nomoto and Yuji Matsumoto. A new approach to unsupervised text summarization. In *Proceedings of the 24th International ACM SIGIR Conference on Research and Development in Informational Retrieval (SIGIR)*, pp. 26–34, 2001.
- [6] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 311–318, 2002.
- [7] 田中英輝, 美野秀弥. やさしい日本語によるニュースの書き換え実験. 情報処理学会研究報告自然言語処理, Vol. 2010-NL-199, No. 11, pp. 1–8, 2010.
- [8] 奥村学, 難波英嗣. テキスト自動要約. オーム社, 2005.
- [9] 西川仁. 自動要約の研究動向: これまでとこれから. 情報処理学会研究報告自然言語処理, Vol. 2013-NL-213, No. 12, pp. 1–1, 2013.
- [10] 梶原智之, 山本和英. 小学生の読解支援に向けた複数の換言知識を併用した語彙平易化と評価. 言語処理学会第 19 回年次大会発表論文集, pp. 272–275, 2013.