Towards Accessible Abstractive Text Summarization

Hacia los resúmenes abstractivos y accesibles

Tatiana Vodolazova
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
tvodolazova@dlsi.ua.es

Abstract: In Natural Language Processing, text summarization and text simplification are the two areas that improve information access for the user. This PhD research project investigates the possibility of integrating text simplification into the abstractive text summarization framework for the purpose of adapting generated summaries to user language proficiency and cognitive ability.
Keywords: Language generation, abstractive text summarization, readability

Resumen: En el Procesamiento de Lenguaje Natural, el resumen y la simplificación de texto son dos áreas cuyo objetivo es mejorar el acceso de la información al usuario. Esta tesis doctoral investga la posibilidad de integrar la simplificación del texto en el marco de generación de resúmenes abstractivos como componente esencial para adaptar resúmenes generados a la competencia lingüística y cognitiva del usuario.
Palabras clave: Generación de lenguaje, resúmenes abstractivos, legibilidad

1 Motivation

The right to information is defined as a basic human right by UNESCO[1], but in the age of data overload accessing the required information is not a straightforward task. The amount of data and their semantic and syntactic complexity require the development of automatic methods capable of representing information in both a compact and comprehensible way. Within the field of natural language processing (NLP), text summarization and text simplification are the two areas of text-to-text generation that can tackle this task.

Text simplification aims to transform complex text into a more comprehensible version while preserving its underlying meaning. The main tasks in text simplification include readability assessment, lexical and syntactic simplification (Saggion, 2017). They encompass a broad range of techniques from designing readability formulas to developing complex word substitution and sentence simplification algorithms.

The main goal of text summarization is to generate a shorter version of the original data while preserving its main concepts, cohesion and grammatical accuracy (Gupta and Gupta, 2018). Text summarization is classified into extractive and abstractive. Extractive approaches produce summaries through selection and concatenation of original text segments. These approaches reveal a number of weaknesses that include:

- concatenation of non-adjacent text segments increases the risk of "dangling anaphora" (i.e. pronouns without referents or with the incorrect ones) and misleading temporal expressions (Steinberger et al., 2007; Smith, Danielsson, and Jönsson, 2012);
- tendency to include lengthy sentences that, apart from the essential information, carry irrelevant text segments (McKeown et al., 2005);
- highly incoherent summaries that fail to convey the gist, especially in the case of concatenating non-adjacent text segments in documents with a high degree of polarized opinions (Cheung, 2008);
- information representation is identical to the original text. In the worst case scenario, where essential knowledge is scattered across all text segments, the generated summary would contain all the original text segments.

[1]https://en.unesco.org/themes/access-information

Lloret, E.; Saquete, E.; Martínez-Barco, P.; Sepúlveda-Torres, R. (eds.) *Proceedings of the Doctoral Symposium of the XXXV International Conference of the Spanish Society for Natural Language Processing (SEPLN 2019)*, p. 63–68
Bilbao, Spain, September 25th 2019.

These deficiencies hinder the extraction of the key concepts and, at the same time, they affect readability of generated summaries making them less comprehensible.

In recent years, interest in text summarization has switched towards abstractive approaches (Gupta and Gupta, 2018). Unlike extractive summarization, abstractive summarization methods aim to generate partially or completely novel text segments. Abstractive text summarization methods that involve natural language generation tools, such as sentence realizers, can generate sentences with resolved agreements (Genest and Lapalme, 2012). As an input, a sentence realizer requires base forms of words and a sentence structure in terms of syntactic constituents. Control over sentence realization addresses the limitations faced by the extractive approach, such that anaphoric expressions and relative information importance are usually resolved on the representation level, while sentence length depends on the chosen sentence structure.

Both text summarization and text simplification are designed to improve information accessibility, but from different perspectives: summarization reduces text volume to the key concepts and simplification makes it more comprehensible.

This PhD research project explores the possibility of integrating text simplification within the framework of abstractive text summarization in order to generate summaries adapted to user language proficiency, knowledge and cognitive ability. This proposal is designed around the hypothesis that abstractive paradigm provides deep control of summarization process that enables the required flexibility to incorporate simplification techniques of both a syntactic and lexical nature. The focus is on examining, applying and analyzing the impact of different techniques and approaches in order to detect the most auspicious ones. Through their optimal combination, the objective is to develop an accessible abstractive text summarization approach.

One of the possible research scenarios focuses on second language (L2) learners who will benefit from the proposed summarization approach. Since the implementation of Common European Framework of Reference for Languages (CEFR) grading scale(Council of Europe, 2001), all texts for L2 learners are graded according to these proficiency guidelines. These texts are written in a clear style and include main communication concepts as well as only the necessary linguistic features corresponding to each linguistic level. They offer a perfect environment for experiments with readability assessment and text summarization.

2 Background and Related Work

Due to the exponential growth of textual data on the web, manual filtering and extraction of necessary information is a tedious and time-consuming task. The first attempt to tackle this task occurred in the mid-twentieth century, when Luhn (1958) designed the first extractive approach to text summarization. Since then, this area of NLP has been extensively researched, exploiting a wide range of both extractive and abstractive techniques (Gupta and Gupta, 2018; Gambhir and Gupta, 2017).

Automatic text simplification, on the other hand, has become an established NLP field only recently. It was designed originally to solve the problem of reduced literacy, but has been also shown to benefit L2 learners, children, people with limited domain knowledge and with cognitive difficulties, such as dyslexia or aphasia (Siddharthan, 2014). Text simplification involves a number of transformations that include sentence split, sentence deletion, insertion, reordering and substitution among others (Saggion, 2017).

To the best of our knowledge there have been very few studies, all conducted by the same authors, that aim to generate accessible summaries through integration of text simplification into summarization process. These authors designed an extractive summarization approach based on a differential evolution algorithm (Nandhini and Balasundaram, 2014). Their method represents each sentence as a set of 4 informativeness features (sentence position, title similarity, etc.) and 5 readability features (word length, sentence length, etc.). Summarization is considered as an optimization problem that aims to maximize both the informativeness and the readability scores. However, their approach is based on the extractive summarization techniques and doesn't involve any simplification. At the same time their set of readability features is small.

3 *Main Hypothesis and Objectives*

This PhD research project explores the possibility of integrating text simplification into the framework of abstractive text summarization in order to generate summaries adapted to user language proficiency, domain knowledge and cognitive ability. It is based on the hypothesis that natural language generation plays a key role for this integration by providing access to and manipulation of both deep semantic and syntactic data structure.

Evaluation of this hypothesis requires research, analysis and development of summarization, natural language generation, simplification and readability assessment techniques with their subsequent application to generate accessible abstractive summaries. To achieve this goal, the following subobjectives are proposed:

- to conduct exhaustive research in text summarization, language generation, text simplification and readability assessment tasks, analyzing current approaches;
- to investigate, propose and analyze new approaches for these tasks and for the intelligent representation of extracted information using techniques based on NLP, focusing on syntactic and semantic knowledge;
- to design the application of the proposed approach for automatic summary generation following an abstractive paradigm;
- to exhaustively evaluate both the proposed approach and the produced summaries. The evaluation will consist of both intrinsic and extrinsic, quantitative and qualitative techniques;
- to analyze possible extensions of the proposed approach for other languages and different user profiles; and,
- to draw conclusions and outline benefits of this research together with a proposal for future work.

4 *Methodology and the proposed experiments*

Since this research encompasses a number of NLP areas, designing an accurate set of experiments requires a clear understanding of where and how these areas interact. For this purpose we need to identify the stages of the process and define the workflow direction. The approach proposed by this PhD research project consists of 6 main stages, namely, information extraction, storage, scoring, text planning, adaptation and text generation. Each of these stages poses a set of corresponding questions that include some of the following:

1. What features to identify in the process of information extraction?
2. How to store extracted information?
3. How to rank each piece of coherent information?
4. How to combine the selected pieces of information?
5. How to assess readability and what kind of simplification techniques to use?
6. How to generate text from the selected information?

Though defined as separate issues, all of these questions are interrelated and cannot be handled in a linear order. For example, depending on the required level of simplification a different piece of duplicated information, based on its readability level, may be selected during the text planning stage.

4.1 Relevant Features

The first set of experiments aims to identify the key features that need to be extracted from the raw text in order to benefit both the process of text summarization and simplification. In our initial research we analyzed whether semantic information such as word senses, anaphora resolution and textual entailment improve informativeness of extractive summaries (Vodolazova et al., 2012). Experiments showed that the combination of the 3 techniques outperforms the baseline and some of the existing summarization systems and, at the same time, it benefits the summarization process more than each technique individually.

This analysis was followed by a closely related experiment that considered whether any type of text can equally benefit from these techniques. The experiment setup involved the evaluation of certain linguistic properties of the original text related to anaphora resolution and textual entailment such as, proper noun, pronoun and noun ratios, and how they affect informativeness of extractive summaries (Vodolazova et al., 2013a). As expected, the results showed that high ratios of at least 2 of these linguistic properties introduce a lot of ambiguity and that the available tools could not handle it. This decrease in the quality of generated summaries emphasized the need for an additional text analysis stage that would help to identify the most favourable summarization technique depending on the linguistic properties of the original text.

The informativeness of generated summaries is not the only goal of our approach. While semantic information, under certain conditions, benefits the informativeness, it may not necessarily benefit the readability. An additional experiment studied how the same semantic techniques within the framework of extractive summarization affected readability of the generated summaries (Lloret et al., 2019). It was shown that, depending on the chosen readability metric, evaluation of informativeness versus readability can generate conflicting results. Out of 8 tested readability metrics, the extractive summarization approach that involves a combination of word sense disambiguation, anaphora resolution and textual entailment scored best only on 3 of the metrics. At the same time, the summarization approach that is based on anaphora resolution and delivered the worst ROUGE(Lin, 2004) results scored best on the other 3 readability metrics.

4.2 Information Representation

Both simplification and abstractive summarization methods require a deep analysis of original data to extract their semantic and syntactic information. This information can be manipulated to generate adapted summaries while maintaining the original meaning and correct grammar. In our initial research within the framework of extractive summarization we experimented with a simplified abstract data representation in a form of a bag of enriched words (Vodolazova et al., 2013b). Each word was either an instance of a function or of a content word, with the latter carrying information about its word sense, concept frequency, part of speech and others.

However, for a fully abstractive summarization approach that uses a sentence realizer for text generation, this representation lacks information about semantic roles, voice, etc. This information will also be required during the readability adjustment stage in order to, for example, convert passive constructions into active ones for the purpose of syntactic simplification. In our first approximation of an abstractive method we designed an abstract representation based on the concept of subject-verb-object triplets. We adapted terminology proposed by Genest and Lapalme (2011) and referred to them as information items (`InIts`). Each `InIt` represents a piece of coherent information. This representation was used to generate ultra-concise summaries within an abstractive summarization method that obtained better results in terms of informativeness than other summarization approaches. The next step would include readability evaluation of summaries generated from this abstract representation.

4.3 Scoring

The scoring stage may be considered as a component of the actual summarization process rather than of the preprocessing stage described so far. The aim of the scoring stage is to rank informativeness of `InIts` with repect to the selected set of features. Our experiments with extractive summarization methods showed that scoring based on concept frequency with resolved anaphoric relations and disambiguated word senses improves the informativeness of summaries (Vodolazova et al., 2012). Similarly, a different experiment with an abstractive summarization approach that scores `InIts` on subject-verb-object and named entities frequencies was shown to outperform other summarization systems (Lloret et al., 2015).

We plan to design the next set of experiments for the scoring stage around the combination of all features that we have tested. Assigning weights to different features according to their impact on informativeness and readability may also benefit the summarization process. Another extension to the scoring stage may involve the integration of readability either as a separate feature or, following the example of Nandhini and Balasundaram (2014), by combining it with the informativeness features in a composite score.

4.4 Text Planning

Once the `InIts` have been scored, it may be sufficient to generate a summary by selecting the top ranked `InIts` individually and converting each one to text until the required summary size has been reached. In this case, the present stage may be omitted. However, mere scoring may be insufficient to produce well-formed summaries. The text planning stage raises a number of challenges that require additional experiments, such as:

- Redundancy detection is the most evident. Redundancy may be present both in terms of identical (or semantically very related) `InIts` and repeated subjects in adjacent sentences;
- Context information, namely, whether the method should select only the highest ranked `InIts` or whether it should include the `InIt` that precedes the highest ranked ones in the original text; and,
- Compression rate considerations, whereby for higher compression rates (i.e. shorter summaries) it may be beneficial to include more `InIts` by reducing noun phrases to their head nouns.

4.5 Readability Assessment

Once provided with the simplification requirement, the readability of each `InIt` needs to be assessed. This may involve conversion of passive constructions into active ones, substitution of long and infrequent words with their shorter and more frequent counterparts.

The exact simplification techniques that may be involved at this stage will be governed by the simplification requirements. The readability metrics that we used in our initial research are of a general nature and all belong to the same family of superficial length-based metrics (Lloret et al., 2019). They neither reflect syntactic nor lexical complexity. Future experiments in the field of readability assessment will include an in-depth research of readability metrics and corresponding simplification techniques for each of the possible target groups, including (but not limited to) L2 learners.

4.6 Text Generation

The final stage of our proposal uses a text realizer to generate sentences from the selected `InIts`. To evaluate the quality of generated sentences we conducted some initial experiments with our first approximation to abstractive summarization. The results showed a decrease in the informativeness of generated summaries when compared to the original sentences (Lloret et al., 2015). However, this method outperformed some other summarization approaches. We will repeat this experiment once all the aforementioned stages are fully developed and integrated with redundancy detection, anaphora resolution, as well the other open issues previously mentioned.

5 *Issues to Discuss*

This paper describes a research proposal that focuses on examining how text summarization and text simplification can be combined in order to make information accessible through adaptation of summaries to the users with different language proficiency levels and cognitive abilities. The outlined approach raises the following issues for discussion:

- A general structure of the approach describing the stages involved and the workflow has been defined. Each stage will trigger a series of experiments in order to analyze and determine its most auspicious implementation. However, is it the optimal combination of stages? Does each stage meet its objective or should it integrate additional functions?
- If we had to choose between syntactic and lexical simplification of summaries, which would be the most appropriate for the L2 learners target group?
- What semantic readability metrics would be the most representative of the target group and what source can be used to gauge their distribution across different levels?

Acknowledgements

This research work has been partially funded by the University of Alicante (Spain), Generalitat Valenciana and the Spanish Government through the projects SIIA (PROMETEU/2018/089), LIVING-LANG (RTI2018-094653-B-C22), INTEGER:(RTI2018-094649-B-I00) and Red iGLN (TIN2017-90773-REDT).

References

Cheung, J. C. 2008. *Comparing abstractive and extractive summarization of evaluative text: controversiality and content selection.* Ph.D. thesis, Department of Computer Science of the Faculty of Science, University of British Columbia.

Council of Europe. 2001. *The Common European Framework of Reference for Languages.* Cambridge University Press.

Gambhir, M. and V. Gupta. 2017. Recent automatic text summarization techniques: A survey. *Artificial Intelligence Review*, 47(1):1–66, jan.

Genest, P.-E. and G. Lapalme. 2011. Framework for abstractive summarization using text-to-text generation. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 64–73, Portland, Oregon, June. Association for Computational Linguistics.

Genest, P.-E. and G. Lapalme. 2012. Fully abstractive approach to guided summarization. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers*, volume 2 of *ACL '12*, pages 354–358, Stroudsburg, PA, USA. Association for Computational Linguistics.

Gupta, S. and S. Gupta. 2018. Abstractive summarization: An overview of the state of the art. *Expert Systems with Applications*, 121:49–65.

Lin, C.-Y. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81, Barcelona, Spain, July. Association for Computational Linguistics.

Lloret, E., E. Boldrini, T. Vodolazova, P. Martínez-Barco, R. Muñoz, and M. Palomar. 2015. A novel concept-level approach for ultra-concise opinion summarization. *Expert Syst. Appl.*, 42(20):7148–7156.

Lloret, E., T. Vodolazova, P. Moreda, R. Muñoz, and M. Palomar. 2019. Are better summaries also easier to understand? analyzing text complexity in automatic summarization: Challenges, models, and approaches. In M. Litvak and N. Vanetik, editors, *Multilingual text analysis challenges, models, and approaches.* World Scientific, New Jersey, 02, pages 337–369.

Luhn, H. P. 1958. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165, April.

McKeown, K., J. Hirschberg, M. Galley, and S. Maskey. 2005. From text to speech summarization. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05).*, volume 5, pages 997–1000. IEEE, March.

Nandhini, K. and S. R. Balasundaram. 2014. Extracting easy to understand summary using differential evolution algorithm. *Swarm and Evolutionary Computation*, 16:19 – 27.

Saggion, H. 2017. Automatic text simplification. *Synthesis Lectures on Human Language Technologies*, 10(1):1–137.

Siddharthan, A. 2014. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298.

Smith, C., H. Danielsson, and A. Jönsson. 2012. A more cohesive summarizer. *Proceedings of COLING 2012: Posters*, pages 1161–1170.

Steinberger, J., M. Poesio, M. A. Kabadjov, and K. Ježek. 2007. Two uses of anaphora resolution in summarization. *Information Processing & Management*, 43(6):1663–1680, November.

Vodolazova, T., E. Lloret, R. Muñoz, and M. Palomar. 2012. A comparative study of the impact of statistical and semantic features in the framework of extractive text summarization. In *Text, Speech and Dialogue - 15th International Conference, TSD 2012, Brno, Czech Republic*, pages 306–313.

Vodolazova, T., E. Lloret, R. Muñoz, and M. Palomar. 2013a. Extractive text summarization: Can we use the same techniques for any text? In E. Métais, F. Meziane, M. Saraee, V. Sugumaran, and S. Vadera, editors, *Natural Language Processing and Information Systems*, pages 164–175, Berlin, Heidelberg. Springer Berlin Heidelberg.

Vodolazova, T., E. Lloret, R. Muñoz, and M. Palomar. 2013b. The role of statistical and semantic features in single-document extractive summarization. *Artificial Intelligence Research*, 2(3):35–44.