POLITECNICO DI TORINO
Repository ISTITUZIONALE

Benchmarking Abstractive Models for Italian Legal News Summarization

Original
Benchmarking Abstractive Models for Italian Legal News Summarization / Benedetto, Irene; Cagliero, Luca; Tarasconi, Francesco; Giacalone, Giuseppe; Bernini, Claudia. - ELETTRONICO. - 379:(2023), pp. 311-316. (Intervento presentato al convegno JURIX2023: 36th International Conference on Legal Knowledge and Information Systems tenutosi a Maastricht (NLD) nel 18-20 December 2023) [10.3233/faia230980].

Availability:
This version is available at: 11583/2987349 since: 2024-03-27T11:06:10Z

Publisher:
IOS PRESS

Published
DOI:10.3233/faia230980

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

21 December 2024

Legal Knowledge and Information Systems
G. Sileno et al. (Eds.)

© 2023 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA230980

Benchmarking Abstractive Models for Italian Legal News Summarization

Irene BENEDETTO [a,b,1] Luca CAGLIERO [b] and Francesco TARASCONI [a] and Giuseppe GIACALONE [c] and Claudia BERNINI [c]
[a] *MAIZE, Via San Quintino, 31, Torino, 10121, Italy*
[b] *Politecnico di Torino, Corso Castelfidardo, 39, Torino, 10129, Italy*
[c] *Giuffrè Francis Lefebvre, Via Busto Arsizio, 40, 20151, Milano, Italy*
ORCiD ID: Irene Benedetto https://orcid.org/0000-0001-7086-7898, Luca Cagliero https://orcid.org/0000-0002-7185-5247

Abstract. Automated text summarization is particularly important in the legal domain due to the length and inherent complexity of the analyzed documents. The Legal AI community has already started to address the text summarization problem. However, most existing approaches focus on English-written documents. Up to now, limited efforts have been devoted to summarizing Italian legal documents. Existing approaches extract portions of existing content without rephrasing them. To bridge this gap, in this work we aim at generating abstractive summaries of Italian legal news. We propose to condense the original news content into different summary types, i.e., an abstract, a title, or a subheader. We benchmark different state-of-the-art summarization models to generate abstractive summaries of Italian legal news. We also investigate the suitability of augmented models capable of handling long Italian documents. The experimental results achieved on a proprietary Italian dataset show the effectiveness of abstractive models in generating fairly accurate summaries and the importance of using larger contextual windows to generate news abstracts.

Keywords. Legal AI, Automated Text Summarization, Abstractive Models

1. Introduction

Legal documents such as legal news, laws, patents, and court judgments are often characterized by large size and complex structure. Legal text is also quite redundant and verbose. These issues pose relevant challenges for content accessibility as document exploration can be extremely time-consuming even for legal experts. Automated text summarization techniques leverage Deep Learning and Natural Language Processing techniques to simplify their access and exploration. Summarizing legal news articles allows legal experts to save both time and resources by gaining a quick insight into the news content. The legal document summarization problem has recently gained the attention of the Legal AI community. For example, in [1] the authors analyze the performance of summarization algorithms' on legal cases and court judgments. As surveyed by [2],

[1] Corresponding Author: Irene Benedetto, irene.benedetto@{polito.it, maize.io}

most existing approaches focus on summarizing English-written legal documents. They encompass summarization approaches suited to various document types among which court judgments [3] and legal case documents [1,4]. More limited efforts are, instead, focused on legal documents written in languages other than English. In the Italian landscape, the most relevant works in the legal domain are ITALIAN-LEGAL-BERT [5] and LamBERTa [6], but they primarily focus on natural language understanding tasks, such as text classification. To the best of our knowledge, the only existing approach that includes Italian legal documents is focused on the legal acts from the European Union law platform [7].

In this work, we address the problem of *abstractive summarization* of *Italian legal news*. An abstractive model automatically generates a summary consisting of arbitrary pieces of text, not necessarily consisting of portions of the original content. Abstractive models can be exploited to generate different summary type among which news abstract, title, or subheader. We benchmark various state-of-the-art abstractive summarization models on a proprietary dataset of Italian legal news articles annotated with humanly generated summaries. Specifically, we examine the performance of various models tailored to the Italian language, e.g., mBART [8], BART-IT [9], IT5 [10] and investigate the effects of the Deep Learning architecture used for model training, the maximum document length supported by the architecture encoder, and the output summary type. The empirical results confirm the effectiveness of the tested summarizers on real news data. To generate abstracts of legal news, we also show the importance of extending the capabilities of existing architectures to handle longer text.

2. Proprietary dataset

The dataset encompasses a wide range of law areas including *Accounting*, *Finance*, *Tax*, *Business*, *Job*, and *Digital*. The number of news documents vary from few tens (e.g., *Digital*) to 1709 (*Tax*). While the number of input tokens used for news summarization are the same, the output length significantly varies across different summary types. *Subheader* is shorter than *Title*, which in turn is shorter than *Abstract* (on average 61 tokens for Abstract, 16 for Title and 7 for Subheader). Abstractive summaries likely contain new terms with respect to the original news documents. *Abstract* introduces a higher percentage of novel n-grams compared to *Subheader* and *Title* (+10% when considering the percentage of novel unigram). This indicates that *Abstracts*' authors are more likely to rephrase the news content to convey the original message. *Abstract* has higher Coverage and Density values (+0.3 of Coverage and +0.8 of Density with respect to Subheader and Title), indicating that they retain a larger portion of the original content while efficiently conveying information. Conversely, *Subheaders* have the lowest Coverage and Density values, probably due to their concise nature. The Compression ratio showcases the reduction in the number of tokens from the input document to the output summary. As expected, *Subheaders* exhibit the highest compression ratio, yielding the maximal compression of the original data.

3. Methodology

We evaluate various established abstractive and extractive summarization models on the proprietary dataset of Italian legal news. For abstractive summarization, we utilize the following state-of-the-art sequence-to-sequence transformer models: *BART-IT* [9], *IT5* [10], and *mBART* [8].

All the aforesaid models struggle with lengthy legal texts due to the quadratic complexity of their self-attention mechanisms. To address this issue, we augment BART-IT with Local Sparse Global Attention [11], which approximates vanilla self-attention in $O(n)$ while maintaining strong performance. For sake of completeness, we also include in the comparison the state-of-the-art model **BERT extractive** [12,13,14], for extractive summarization, that uses the Italian legal BERT [6].

4. Experiments

Evaluation metrics We adopt the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [15] and BERTScore [16] metrics to evaluate the syntactic and semantic similarity of the outputs with the humanly generated versions.

Experimental settings Experiments were run on a machine equipped with Intel® Core™ i9-10980XE CPU, 1 × Nvidia® V100 GPU, 32 GB of RAM running Ubuntu 22.04 LTS. The choice of hyperparameters is guided by the results on the validation set, and based on a combination of manual and grid search. We test the following the learning rates: 5e-5, 1e-5, 1e-6. All models are trained for a maximum of 10 epochs to maximize log-likelihood through the use of the sequence cross entropy loss and the AdamW optimizer [17] with a weight decay of 0.01. The batch size is tailored based on the model size and hardware capacity and ranges from 8 to 32. The training set consists of the 70% of the proprietary dataset. For model evaluation during training, we use a 10% of the data as a validation set. We select the best checkpoint based on the ROUGE-1 score on the validation set. Finally, the test set consists of the remaining 20% of the data.

4.1. Results

Comparison between summarization models Table 1 presents a performance comparison between different abstractive models, where we separately report the results obtained for different summary types. The performance scores of extractive models are, in general, worse than that of abstractive approaches. Thanks to its larger contextual window, LSG-BART-It consistently outperforms all the other approaches. Attending to larger portions of the input text has shown to be particularly relevant to achieve high-quality summarization results on legal news documents. Among the other approaches, as expected, the large version of IT5 outperforms its base and small versions while BART-IT performs better than both mBART-large and IT5.

The performance of extractive models are, in general, worse than that of abstractive approaches. BERTExtr turns out to be the best-performing extractive approach.

Table 1. Comparison of abstractive models' performance on test data along with the baseline extractive in gray. LSG-BART-It consistently achieves the highest F1-Scores in most tasks.

Target document	Model	ROUGE-1	ROUGE-2	ROUGE-L	BERT-score
Abstract	Baseline - BERTExtr	0.31	0.156	0.224	0.718
	IT5-small	0.368	0.136	0.335	0.559
	IT5-base	0.390	0.159	0.358	0.572
	IT5-large	0.400	0.166	0.368	0.573
	BART-It	0.441	0.231	0.411	0.625
	mBART-large	0.433	0.221	0.395	**0.633**
	LSG-BART-It	**0.456**	**0.249**	**0.427**	0.631
Subheader	Baseline - BERTExtr	0.066	0.028	0.063	0.621
	IT5-small	0.231	0.013	0.229	0.528
	IT5-base	0.228	0.007	0.228	0.450
	IT5-large	0.239	0.016	0.239	0.505
	BART-It	0.313	0.128	0.313	0.642
	mBART-large	0.222	0.061	0.218	0.602
	LSG-BART-It	**0.320**	**0.136**	**0.319**	**0.645**
Title	Baseline - BERTExtr	0.186	0.080	0.150	0.679
	IT5-small	0.245	0.003	0.232	0.541
	IT5-base	0.270	0.021	0.258	0.550
	IT5-large	0.296	0.034	0.282	0.555
	BART-It	0.292	0.047	0.285	0.595
	mBART-large	0.265	0.035	0.256	0.591
	LSG-BART-It	**0.298**	**0.051**	**0.290**	**0.597**

Qualitative examples Table 2 reports some representative examples of summary outcomes. For the sake of readability, we report both the generated and expected summaries written in Italian and its English translation. The achieved outcomes generally exhibits a commendable alignment with the expected outcome, closely matching the expected content and intent of the reference summary. Overall, the results demonstrate the ability of abstractive models to effectively comprehend and synthesize legal content.

5. Conclusion and future work

In this work, we benchmarked abstractive summarization models on Italian legal news documents. It provides legal experts with actionable tools for condensing news content into a more succinct version, e.g., a title or an abstract. The work advances the state-of-the-art work in Italian legal document summarization by leveraging generative models beyond extractive ones. Due to the complexity and length of legal news, augmenting standard models with more advanced capabilities to handle long documents has shown to improve abstractive summarization performance.

Our future research line will encompass the following activities: (1) Specialize the pre-trained models in the legal domain to further improve legal language understanding; (2) Analyze extended versions of the proprietary datasets including additional law areas and languages; (3) Capture nuances of legal reasoning within summaries.

Table 2.: Qualitative examples (LSG-BART-IT vs. reference summary).

Type	Output summary	Reference summary
Abstract	Al fine di contenere gli effetti dell'aumento dei prezzi e delle tariffe del settore energetico per le imprese e i consumatori, la Legge di Bilancio 2023 art. 1 comma -121 L. 197/2022 ha istituito per l'anno 2023 un contributo di solidarietà temporaneo sugli extraprofitti.	Per fronteggiare gli effetti dell'aumento del costi delle tariffe di gas, energia, per imprese e lavoratori, la Legge di Bilancio 2023 ha. 1 comma -121 L. 197/2022 ha previsto per l'anno 2023 un contributo di solidarietà.
Title	Il contributo di solidarietà per il caro energia	Istituito il contributo di solidarietà per il caro energia
Subheading	Nella Legge di Bilancio 2023	Nella Legge di Bilancio 2023
Abstract	In order to contain the effects of rising prices and tariffs in the energy sector for companies and consumers, the 2023 Budget Law art. 1 paragraph -121 L. 197/2022 has established a temporary solidarity contribution on windfall profits for the year 2023.	To address the impact of rising gas and energy tariff costs on companies and workers, the 2023 Budget Law art. 1 paragraph -121 L. 197/2022 has provided for a solidarity contribution for the year 2023.
Title	The Solidarity Contribution for Energy Cost Increases	Establishment of the Solidarity Contribution for Energy Cost Increases
Subheading	In the 2023 Budget Law	In the 2023 Budget Law
Abstract	L'Agenzia delle Entrate, con il provvedimento 116259 del 3 aprile 2023, ha comunicato che la percentuale di fruizione del credito per l'acqua potabile è pari al 17,9005%.	L'Agenzia delle Entrate, con il provvedimento n. 116259 del 3 aprile 2023, ha comunicato che la percentuale di fruizione del bonus acqua potabile è pari al 17,9005%.
Title	Bonus acqua potabile: la percentuale di fruizione	Bonus acqua potabile: definita la percentuale di fruizione
Subheading	Crediti d'imposta	Crediti d'imposta
Abstract	The Revenue Agency, with provision 116259 dated April 3, 2023, communicated that the utilization rate for the drinking water credit is 17.9005%.	The Revenue Agency, with provision no. 116259 dated April 3, 2023, communicated that the utilization rate for the drinking water bonus is 17.9005%.
Title	Drinking Water Bonus: Utilization Rate	Drinking Water Bonus: Utilization Rate Defined
Subheading	Tax Credits	Tax Credits
Abstract	Le domande per le agevolazioni ZFU Sisma Centro Italia possono essere inviate fino alle ore 12.00 di mercoledì 24 maggio 2023. Ogni soggetto interessato può presentare una sola domanda e l'ordine temporale di presentazione delle domande è irrilevante ai fini dell'accesso alle agevolazioni.	Entro le ore 12.00 di mercoledì 24 maggio 2023 le domande per le agevolazioni ZFU Sisma Centro Italia possono essere inviate. Si può presentare una sola domanda e l'ordine temporale di presentazione delle domande è irrilevante ai fini dell'accesso alle agevolazioni.
Title	ZFU Italia: domande entro il 16 maggio 2023	ZFU Sisma Centro Italia
Subheading	Urbana	Zona Franca Urbana
Abstract	Applications for the ZFU Central Italy Earthquake incentives can be submitted until 12:00 PM on Wednesday, May 24, 2023. Each interested party can submit only one application, and the chronological order of application submission is irrelevant for accessing the incentives.	Applications for ZFU Central Italy Earthquake incentives can be submitted by 12:00 PM on Wednesday, May 24, 2023. Only one application can be submitted, and the chronological order of application submission is irrelevant for accessing the incentives.
Title	ZFU Italy: Applications by May 16, 2023	ZFU Central Italy Earthquake Area
Subheading	Urban	Urban Free Zone

References

[1] Shukla A, Bhattacharya P, Poddar S, Mukherjee R, Ghosh K, Goyal P, et al. Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation. In: The 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing; 2022. .

[2] Jain D, Borah MD, Biswas A. Summarization of legal documents: Where are we now and the way forward. Computer Science Review. 2021;40:100388. Available from: https://www.sciencedirect.com/science/article/pii/S1574013721000289.

[3] Bhattacharya P, Hiware K, Rajgaria S, Pochhi N, Ghosh K, Ghosh S. A comparative study of summarization algorithms applied to legal case judgments. In: European Conference on Information Retrieval. Springer; 2019. p. 413-28.

[4] Polsley S, Jhunjhunwala P, Huang R. CaseSummarizer: A System for Automated Summarization of Legal Texts. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations. Osaka, Japan: The COLING 2016 Organizing Committee; 2016. p. 258-62. Available from: https://aclanthology.org/C16-2054.

[5] Licari D, Comandè G. ITALIAN-LEGAL-BERT: A Pre-trained Transformer Language Model for Italian Law. In: Symeonidou D, Yu R, Ceolin D, Poveda-Villalón M, Audrito D, Caro LD, et al., editors. Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management. vol. 3256 of CEUR Workshop Proceedings. Bozen-Bolzano, Italy: CEUR; 2022. ISSN: 1613-0073. Available from: https://ceur-ws.org/Vol-3256/#km4law3.

[6] Tagarelli A, Simeri A. Unsupervised Law Article Mining based on Deep Pre-Trained Language Representation Models with Application to the Italian Civil Code. CoRR. 2021;abs/2112.03033. Available from: https://arxiv.org/abs/2112.03033.

[7] Aumiller D, Chouhan A, Gertz M. EUR-Lex-Sum: A Multi- and Cross-lingual Dataset for Long-form Summarization in the Legal Domain. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics; 2022. p. 7626-39. Available from: https://aclanthology.org/2022.emnlp-main.519.

[8] Liu Y, Gu J, Goyal N, Li X, Edunov S, Ghazvininejad M, et al. Multilingual Denoising Pre-training for Neural Machine Translation. Transactions of the Association for Computational Linguistics. 2020;8:726-42. Available from: https://aclanthology.org/2020.tacl-1.47.

[9] Quatra L, Cagliero. BART-IT: An Efficient Sequence-to-Sequence Model for Italian Text Summarization. Future Internet. 2022 Dec;15(1):15. Available from: http://dx.doi.org/10.3390/fi15010015.

[10] Sarti G, Nissim M. IT5: Large-scale Text-to-text Pretraining for Italian Language Understanding and Generation. ArXiv preprint 220303759. 2022 mar. Available from: https://arxiv.org/abs/2203.03759.

[11] Condevaux C, Harispe S. LSG Attention: Extrapolation of pretrained Transformers to long sequences. In: Advances in Knowledge Discovery and Data Mining: 27th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2023, Osaka, Japan, May 25–28, 2023, Proceedings, Part I. Springer; 2023. p. 443-54.

[12] Liu Y, Lapata M. Text Summarization with Pretrained Encoders. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics; 2019. p. 3730-40. Available from: https://aclanthology.org/D19-1387.

[13] Miller D. Leveraging BERT for Extractive Text Summarization on Lectures. CoRR. 2019;abs/1906.04165. Available from: http://arxiv.org/abs/1906.04165.

[14] Roush A, Balaji A. DebateSum: A large-scale argument mining and summarization dataset. In: Proceedings of the 7th Workshop on Argument Mining. Online: Association for Computational Linguistics; 2020. p. 1-7. Available from: https://aclanthology.org/2020.argmining-1.1.

[15] Lin CY. ROUGE: A Package for Automatic Evaluation of Summaries. In: Text Summarization Branches Out. Barcelona, Spain: Association for Computational Linguistics; 2004. p. 74-81. Available from: https://aclanthology.org/W04-1013.

[16] Zhang T, Kishore V, Wu F, Weinberger KQ, Artzi Y. BERTScore: Evaluating Text Generation with BERT. CoRR. 2019;abs/1904.09675. Available from: http://arxiv.org/abs/1904.09675.

[17] Loshchilov I, Hutter F. Fixing Weight Decay Regularization in Adam. CoRR. 2017;abs/1711.05101. Available from: http://arxiv.org/abs/1711.05101.

Acknowledgements

The authors would like to thank Giuffrè Francis Lefebvre S.p.A. for providing us with the proprietary data. This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from Next-GenerationEU (Italian PNRR – M4 C2, Invest 1.3 – D.D. 1551.11-10-2022, PE00000004). This study was also partially carried out within the FAIR (Future Artificial Intelligence Research) and received funding from Next-GenerationEU (Italian PNRR – M4 C2, Invest 1.3 – D.D. 1555.11-10-2022, PE00000013). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.