ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2026-11-17

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Zanatti, M., Ribeiro, R. & Pinto, H. S. (2025). Segmentation model for judgments of the Portuguese supreme court of justice. In Manuel Filipe Santos, José Machado, Paulo Novais, Paulo Cortez, Pedro Miguel Moreira (Ed.), Progress in artificial intelligence: EPIA 2024 (pp. 245?257). Springer. https://doi.org/10.1007/978-3-031-73497-7_20

Further information on publisher's website:
10.1007/978-3-031-73497-7_20

Publisher's copyright statement:
This is the peer reviewed version of the following article: Zanatti, M., Ribeiro, R. & Pinto, H. S. (2025). Segmentation model for judgments of the Portuguese supreme court of justice. In Manuel Filipe Santos, José Machado, Paulo Novais, Paulo Cortez, Pedro Miguel Moreira (Ed.), Progress in artificial intelligence: EPIA 2024 (pp. 245?257). Springer. https://doi.org/10.1007/978-3-031-73497-7_20, which has been published in final form at https://dx.doi.org/10.1007/978-3-031-73497-7_20. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

Segmentation Model for Judgments of the Portuguese Supreme Court of Justice

Martim Zanatti[1,3][0000−0002−1144−4104], Ricardo Ribeiro[2,3][771D−DE47−EB63], and H. Sofia Pinto[1,3][0000−0001−9209−1349]

[1] Instituto Superior Técnico - Universidade de Lisboa, Av. Rovisco Pais, Portugal
[2] Iscte - Instituto Universitário de Lisboa, Avenida das Forças Armadas, Portugal
[3] INESC-ID, Rua Alves Redol, 9, Lisbon, Portugal

Abstract. Legal document segmentation is a critical task in the field of natural language processing (NLP), enabling efficient analysis, retrieval, and understanding of legal content. Despite its importance, research in this area for European Portuguese has been limited. To address this gap, we present a novel approach to automate the segmentation of legal judgments from the Portuguese Supreme Court of Justice into distinct sections. Leveraging a Bi-LSTM-CRF model, we developed a dataset and achieved significant results, including an accuracy of 0.9997, precision of 0.9986, recall of 0.996, and F1-Score of 0.9973. Our methodology and experimental results demonstrate the effectiveness and potential applications of our approach for the European Portuguese language.

Keywords: Sequence Labeling · Legal Documents Segmentation · Portuguese Legal Documents · Bi-LSTM-CRF · Dataset creation

1 Introduction

Legal documents play a pivotal role as primary sources of information for legal proceedings, analyses, and decision-making across diverse legal and business contexts. However, owing to their extensive length and intricate nature, these documents present formidable challenges for automated analysis and text processing. The capacity to extract pertinent information and meaningful insights from legal documents is indispensable for legal practitioners, scholars, and professionals in related fields. Moreover, ensuring efficient accessibility to legal data is imperative for fostering transparency, regulatory adherence, and informed decision-making.

A fundamental approach to dealing with the complexity of legal documents is segmentation, which involves dividing these documents into smaller, more manageable units. The ability to segment legal documents according to various criteria, such as rhetorical roles and legal sections, holds significant importance for enhancing multiple natural language processing (NLP) tasks. This segmentation aids in reducing the complexity and size of legal documents, addressing a common challenge when employing deep learning models. Moreover, it enables focused attention on crucial information pertinent to specific tasks, thereby enhancing the efficiency and effectiveness of NLP applications in the

legal domain. Segmentation of legal documents can facilitate analysis, research, and information retrieval, enabling a quicker and more effective understanding of legal content. In this context, our work focuses on developing a legal document segmentation model aimed at automating the process of segmenting judgments of the Portuguese Supreme Court of Justice into compact and logical sections, enhancing accessibility to European Portuguese legal data.

Research in the legal domain for European Portuguese has been relatively limited. To address this gap, as part of the *IRIS* project between INESC-ID and the Portuguese Supreme Court of Justice[4] (STJ), we have developed a dataset for segmenting legal judgments into distinct sections from scratch. Leveraging the Bi-LSTM-CRF model, which achieves an **accuracy** of **0.9997**, a **precision** of **0.9986**, a **recall** of **0.996**, and an **F1-Score** of **0.9973**, we aim to provide a robust framework for automated segmentation of Portuguese Supreme Court of Justice (STJ) judgments. In this paper, we present our methodology, dataset, and experimental results, shedding light on the effectiveness and potential applications of our approach in the domain of legal document analysis for the European Portuguese language.

The remainder of the paper is organized as follows. Section 2 addresses what has been achieved in this area over the years. Section 3 introduces the context of our problem, particularly the definition of the sections used for segmenting the judgments. Section 4 explains how our corpus was created. Section 5 describes our method, where we address the process of creating Bi-LSTM-CRF model to our context. Section 6 presents and discusses the results of our system. Section 7 concludes and discusses future work.

2 Related Work

The task of segmenting legal documents into sections, which delineate the logical structure of the document, can be viewed as a sequence labeling problem. By capitalizing on the sequential nature of textual data, models are capable of discerning patterns and dependencies that signify transitions between sections. This facilitates the effective segmentation of documents into their logical structures. Over the years, various approaches have been proposed across different fields to tackle similar segmentation tasks [9].

2.1 Sequence Tagging

The sequence tagging problem, encompassing tasks such as part-of-speech tagging, chunking, and named entity recognition, has remained a cornerstone in natural language processing (NLP) research over the years. This problem has been addressed by statistical models such as Hidden Markov Model (HMM) [8,15,17,18], Conditional Random Fields (CRF) [12,13], and Maximum Entropy models [6,19]. Masked Conditional Random Field, as implemented by [22], imposes restrictions on candidate tagging during both the training and decoding

[4] https://www.stj.pt/

phases to improve consistence in sequence tagging. Convolutional models have also been proposed to cover the sequence tagging problem [5,24].

The challenge of sequence entity tagging was addressed by implementing a Bidirectional Long Short-Term Memory (LSTM) Conditional Random Field (CRF) model, as detailed in [10]. This model integrates both forward and backward passes of the LSTM, allowing for the incorporation of contextual information from preceding and subsequent tokens in the sequence. During the forward pass, the model computes output scores $f_\theta([x]_1^T)$ for all possible tags at each position in the sequence. Subsequently, a CRF layer is applied to conduct forward and backward passes, facilitating the calculation of gradients for both the network output and the state transition edges. This enables the model to capture dependencies between adjacent labels and refine its predictions based on the global context of the entire sequence. Following the forward and backward passes, error backpropagation is performed to update the network parameters, involving the propagation of errors from the output layer back to the input layer, encompassing both the forward and backward states of the LSTM. Consequently, adjustments are made to the network parameters, including the state transition matrix $[A]_{i,j}$ and the original bidirectional LSTM parameters θ. This approach harnesses the strengths of Bidirectional LSTM and CRF models to effectively capture long-range dependencies and sequential patterns within the input data, thereby enhancing the accuracy of sequence entity tagging tasks.

2.2 Email Zoning

In addition to conventional sequence labeling tasks such as part-of-speech tagging and named entity recognition, another pertinent application area is email zoning, which entails segmenting emails into meaningful sections. Unlike part-of-speech tagging, which operates at the word level, email zoning works at the sentence level, with the objective of identifying and categorizing various components of an email, including the salutation, body, and signature. Pioneering work in the field of email segmentation was conducted by [3], who leveraged linguistic and geometrical patterns to detect email signatures. Furthermore, [2] developed JANGADA, a supervised learning system that employed Conditional Random Fields (CRF) and sequence-aware perceptron algorithms to identify signature blocks and quoted text within emails. Additionally, [7] were among the first to introduce a segmentation schema for email text, presenting a corpus of 9,836 recruited respondents' email messages with a broader annotation schema focusing on various email components. The OKAPI system [11] is the first end-to-end multilingual system with pre-trained transformer models for performing email zoning. To address the multilingual nature of emails, a multilingual sentence encoder was developed. XLM-RoBERTa was used to produce embeddings for each token in an email line, and a segmentation model complements the multilingual sentence encoder. The segmentation model is responsible for segmenting the email text into functional zones and is composed of a BiLSTM with a CRF on top to classify each sentence into an email zone.

2.3 Legal Documents Segmentation

Legal documents, known for their extensive length, intricate nature, and structured format, constitute a vital source of information crucial for legal deliberation and decision-making [21]. However, the difficulty in understanding their structure composition introduces obstacles to automated processing and analysis. To enhance the extraction of meaningful insights and facilitate accessibility, diverse methodologies for segmenting legal documents have emerged. Legal document segmentation has been approached from various angles, each addressing specific aspects of the document's content and organization. For instance, [4] proposed a Deep Gating Network designed to predict prison terms by decomposing them into a series of charge-based prison term predictions (CPTPs). Similarly, [23] introduced a label-conditioned Seq2Seq model to address the challenge of Court View generation from fact descriptions in criminal cases.

Similarly to our study, [20] segmented U.S. court decisions, using CRF with handcrafted features, into functional and issue-specific parts. Also, [1] employed a BiLSTM-CRF model with sent2vec features to label 8 rhetorical roles in Indian Supreme Court documents, while [14] extended it by developing a multitask learning-based model to label 13 rhetorical roles for Indian legal documents. The multitask learning framework consists of a rhetorical role predictor comprising a BiLSTM-CRF model, along with a label shift predictor component serving as an auxiliary task. The intuition behind the label shift prediction is derived from the observation that rhetorical role labels exhibit gradual shifts across consecutive sentences, maintaining topical coherence.

3 Context of the Problem

Judgments of the Portuguese Supreme Court of Justice can be divided into several sections, each playing a crucial role in understanding the case. Apart from the three primary sections that consistently appear in judgments – namely, Report (referred to as *relatório* in Portuguese), Ground (referred to as *fundamentação* in Portuguese) that is divided into Facts (*fundamentação de facto*) and Law (*fundamentação de direito*), and the Decision section – there can be six additional sections: **Header** (*cabeçalho*), **Delimitation** (*delimitação*), **Signature** (*colectivo*), **Declaration** (*declaração*), **Footnotes** and **Title**.

The **Header** section serves as the introductory part of the judgment, comprising crucial elements such as **Document Identification**, which designates the document as a legal judgment from a particular court. The **Case Reference Number** serves as a unique identifier for the judgment or case, essential for record-keeping and future referencing. Additionally, the **Type of Procedure** indicates the nature of the case or legal procedure, categorized into criminal, civil, administrative, among others. The **Report** section introduces the content of the judgment, what the various concerned sides are seeking and which decision is to be considered and decided. The **Grounds** section, is usually divided into two sub-sections: **facts** and **law**. The first sub-section sets out the proven facts of the case while the second sub-section argues about everything that has been

elaborated so far to make the final decision. The **Decision** is a small section, only containing the final decision. The **Delimitation** section serves a crucial purpose in framing the main content of the judgment. It outlines the specific matters the court intends to address, defines the scope of the case, elucidates primary legal questions, and identifies any specific issues or concerns the court will evaluate. The **Signature** section lists the names of the three judges who contributed to the case and it includes the date as well. In the **Declaration** section, judges provide official statements or clarifications regarding their decision. This includes affirming the court's decision, presenting dissenting opinions, and offering clarifications to provide additional insight into their reasoning. **Footnotes** in legal judgments play a crucial role, diverging from traditional usage found in documents like scientific papers. Extensive and ubiquitous throughout judgments, footnotes serve various key functions. They include **Legal Citations**, primarily citing statutes, regulations, case law, and other legal sources, forming the basis of legal arguments. **Clarification**, providing additional context or insights without interrupting the main narrative. **Historical Context**, offering background on legal principles, past decisions, or statutes. **Additional References**, directing readers to secondary sources for broader understanding or different perspectives. **Cross-referencing**, aiding navigation through complex arguments or discussions within lengthy judgments.

Within the structural framework of Portuguese judgments, the initial segment is usually recognized as the **Header** section. This section is not a universal feature across all judgments issued by the Portuguese Supreme Court of Justice. Following this, the **Report** section consistently presents itself in every judgment. In certain judgments, a **Delimitation** section can be found nestled between the Report and the **Grounds** sections. The latter is usually divided into **facts** and **law**, where the law section is an omnipresent feature in judgments. The **Decision** section succeeds the Grounds section and is another omnipresent element. Subsequent sections include the **Signature** and **Declaration** sections, however, they may not be present and their order can vary. In certain judgments, **Titles** are assigned to specific sections, and **Footnotes** are also present in some cases.

4 Dataset

To effectively train our Segmenting Judgment Model, a dataset segmented into the specified sections is crucial. However, since there was none for the Portuguese Supreme Court of Justice judgments, the creation of a dataset from scratch was required. This was done semi-automatically: sections beginning with a title enabled automated segmentation, followed by a script to verify text-section associations. Any discrepancies were manually corrected. We constructed a dataset comprising 670 distinct judgments of the Portuguese Supreme Court of Justice. Although our architecture was trained using these judgments, it became evident that they all adhered to a similar structure. Consequently, the model underperforms when presented with judgments that deviate from this structure. To address this issue, we created variations of the original 670 judgments by omit-

ting specific sections that were not ubiquitous in all judgments, namely: Header, Delimitation, Signature, Declaration, Footnotes, and Titles. This yielded a diversified dataset of 11,904 judgments, containing various structural variations.

Given that the dataset was not curated by legal experts, we rely on the titles within the judgments to partition them into sections. Nonetheless, certain judgments lack explicit titles, necessitating manual separation. To address the inherent challenges associated with manual segmentation by non-experts, we ensured that the test dataset exclusively comprised judgments already divided into sections by the respective titles authored by the judges themselves. This approach ensured that our results aligned with the divisions made by the judges, enhancing the credibility and reliability of our segmentation model.

5 Method

For training the Segmenting Judgment Model we adopted a Bidirectional LSTM-CRF model, which is similar to the one used by [10]. The Bidirectional LSTM component is designed to capture sequential information from both forward and backward directions. This capability is particularly useful for sequence tagging tasks, where input features from both the past and future are available for a given point in time. The Conditional Random Field (CRF) is a probabilistic graphical model specifically designed for sequence labeling tasks. When integrated with a Bidirectional LSTM, the subsequent CRF layer leverages the representations produced by the Bidirectional LSTM to deliver globally normalized sequence predictions, thereby ensuring the coherence of the predicted tag sequence.

Adapting Zhiheng Huang et al. [10] foundational work to our problem required critical modifications to align with our objectives. Their model, focusing on granular sequence tagging at the sentence level, was tailored for individual words. In contrast, we broadened our scope to analyze judgments from the Portuguese Supreme Court of Justice. Rather than parsing sentences, we examined entire paragraphs, assigning each a section tag. While sentences rely on neighboring words for context, paragraphs express complete thoughts within the judgment's narrative structure. Understanding this hierarchy is crucial for accurate segmentation. This adaptation addressed the need to systematically structure Portuguese legal documents, enhancing accessibility and practicality. Unlike the original model, which used padding to maintain consistent sentence length, we recalibrated at the document level, ensuring all judgments had an equal number of paragraphs. This adjustment prevented inconsistencies during processing, considering the variability in paragraph count within legal judgments. Furthermore, while the original model employed named entity tags, our adaptation tailored tags to Portuguese judgments' context. Instead of generic tags, our model recognizes and assigns section names specific to judgments, as explained in Section 3. Despite this semantic shift, both approaches aim to capture essential text elements by assigning specific tags to words or paragraphs. This contextual tagging was essential for our model's applicability and effectiveness in analyzing Portuguese legal judgments. Figure 1 depicts the architecture of the Segment-

ing Judgment Model. Our method aligns with the approach by Zhiheng Huang et al. In this structure, each section (excluding the title tag) is divided into a B-section and an I-section, where the B-section denotes the beginning of the respective section, while the I-section denotes its intermediate position.

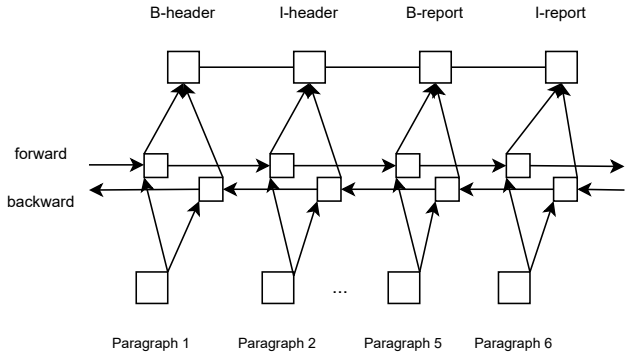


Fig. 1. Segmenting Judgment Architecture. Here, the transition from the Header section to the Report section is illustrated.

5.1 Training Procedure

We partitioned the 11,904 judgments into training (with a portion for cross-validation) and test sets: 80% was allocated for both training and cross-validation, leaving 20% for testing. During the cross-validation and hyper-parameter optimization process, we used 10 distinct folds, shuffling the data to ensure a robust evaluation across diverse data subsets. The best-performing hyper-parameters are as follows: 10 training epochs per fold with a batch size of 46. This configuration allows us to process 46 judgments concurrently without limiting the number of paragraphs in each. We used the Adam optimizer, set to a learning rate of 9.835×10^{-4}, and the network comprises one hidden layer with 971 units and a dropout rate of 0.44. Our early stopping mechanism operates on a "patience" threshold of 4: if there is no decrease in loss over an epoch, the "patience" increments. Upon reaching a "patience" of 4, a "decay" counter increases. The training halts prematurely when the "decay" counter reaches 4. We employ the *stjiris/bert-large-portuguese-cased-legal-mlm-nli-sts-v1* model[5] [16] to encode paragraphs into embeddings. This model, a legal adaptation of BERTimbau large, represents each paragraph in a 1024-dimensional dense vector space.

[5] https://huggingface.co/stjiris/bert-large-portuguese-cased-legal-mlm-nli-sts-v1

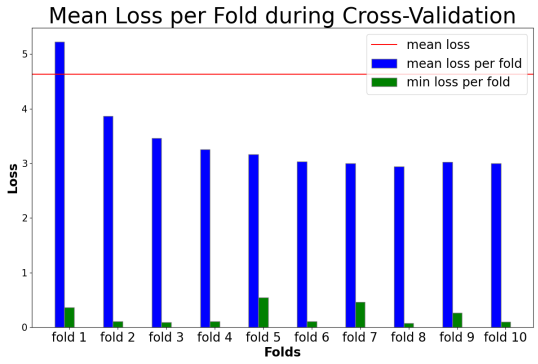


Fig. 2. Graphical representation of the loss distribution across 10 folds: showcasing the mean loss for each individual fold, the collective mean loss for all folds, and the minimum loss recorded for every fold.

In Figures 2 and 3 we can observe that the average loss across the folds consistently ranges between 2.9 and 5.2, with an overall mean of 4.6, indicating model stability across different data subsets. However, the stark contrast between the average losses and the considerably lower minimum loss values, which range between 0.08 and 0.5, suggests moments of optimal model fitting to certain data segments. This wide variance might be attributed to the use of a 0.44 dropout rate during training, which could cause initial high deviation losses, starting above 35, and then sharply reducing to the aforementioned minimums. The similar learning curves among folds signify model consistency. The model's performance on the test set, which encompasses 20% of the dataset, offers a promising counterpoint to potential overfitting concerns.

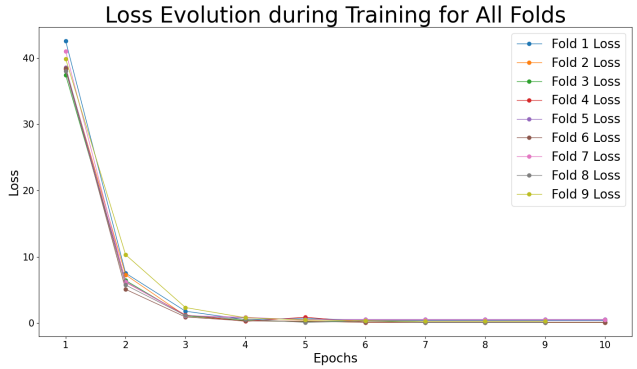


Fig. 3. Loss evolution across 10 epochs for each of the 10 cross-validation folds.

6 Results

The model exhibits an **accuracy** of **0.9997**, which showcases its ability at making correct predictions. A **precision** of **0.9986**, indicates that the majority of its positive predictions are indeed correct. A **recall** of **0.996** suggests it successfully identifies most of the positive instances in the dataset. An **F1-Score** of **0.9973**, which harmoniously balances precision and recall, further enhances the model's robustness. These results paired with the consistent behavior observed across all cross-validation folds, bolster the claim of strong generalization capabilities.

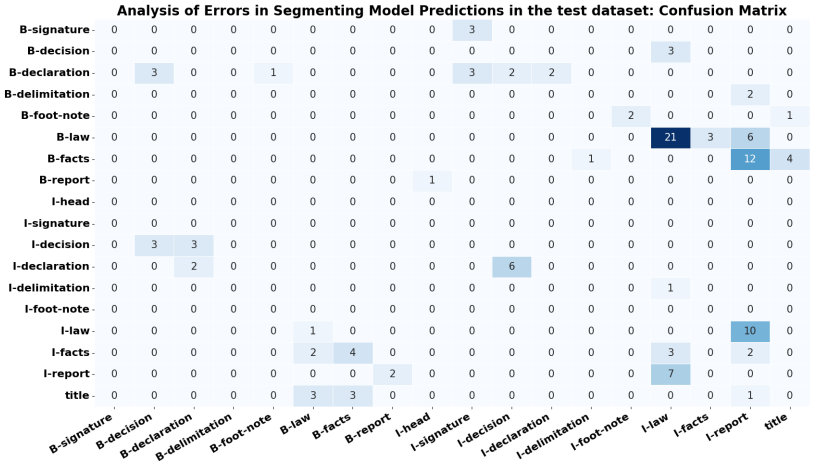


Fig. 4. Confusion matrix providing an in-depth analysis of errors made by the Segmenting model in predicting judgment sections. In each cell is the frequency of misclassifications, showing only the errors made by the model. For instance, the cell at row "B-facts" and column "I-report" indicate the number of times the model incorrectly predicted the "I-report" instead of "B-facts".

Among the 457,349 paragraphs in the test dataset, the model misclassified 123 paragraphs. Figure 4 illustrates the nature of errors attributed to the segmentation model. Notably, with all sections except the title section composed of "B-section" and "I-section" labels, two distinct types of errors emerge. First, misclasifications arise when the model correctly identifies the section but misinterprets its position as either the beginning or intermediate. Secondly, more significant errors occur when the model incorrectly classifies a paragraph into the wrong section altogether. Excluding the first type of error, the number of misclassifications is 102. In fact, as illustrated in Figure 4, the most common error falls in the first category, between "B-law" and "I-law" sections.

We hypothesize that the presence of titles separating "report", "facts", "law", and "decision" sections from other sections contributed to the high results obtained. To mitigate this potential "bonus" feature, we included judgments without any titles in the training, validation, and test datasets. However, to assess the model's adaptability to judgments without titles, we removed every title from the test dataset before evaluation. Without any titles present in the test dataset, we achieved an accuracy of **0.997**, a precision of **0.9458**, a recall of **0.94**, and an F1-score of **0.9428**. Although these results are slightly lower than those obtained with titles, they still show a very high level of performance.

Among the 449,031 paragraphs in this test dataset, the model misclassified 157 paragraphs. The total number of errors in the second category is 106. As depicted in Figure 5, the most common mistake occurs between "B-law" and "I-law" labels. As expected, since the title section is not present, this error appears more frequently in this dataset. This is because the title section marks the beginning of a new section in the next paragraph. Another error that increases, for the same reason, is the misclassification between "B-facts" and "I-report" labels. With the absence of the title section, since the facts section follows the report section, the model faces greater difficulty in discerning section changes.

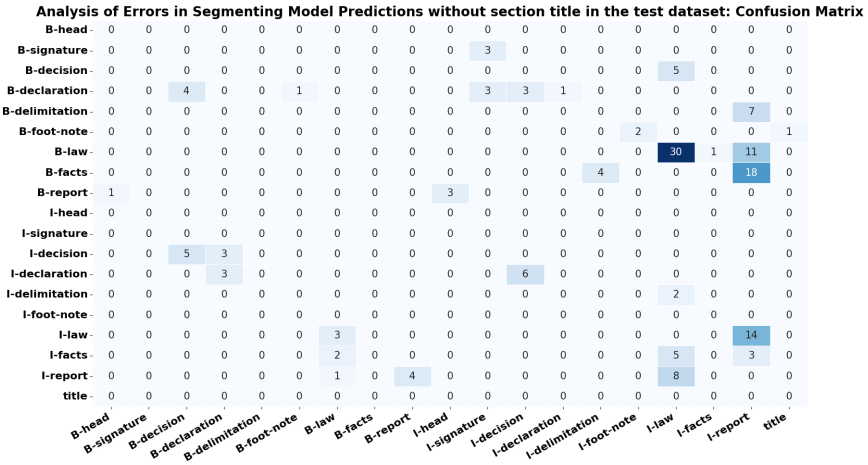


Fig. 5. Confusion matrix providing an in-depth analysis of errors made by the Segmenting model in predicting judgment sections after removing the section title from all judgments. In each cell is the frequency of misclassifications, showing only the errors made by the model.

7 Conclusions and Future Work

In this paper, we introduce a segmentation model for judgments of the Portuguese Supreme Court of Justice. Our efforts included the creation of a dataset

from scratch and the training of the Segmenting Judgment Model to partition Portuguese Supreme Court of Justice judgments into sections, thereby enhancing our understanding of the structural aspects of these judgments. Structuring judgments into sections can indeed offer significant benefits for various tasks. Judges, for example, can easily locate specific sections within the judgment, thereby reducing search time, which can often be expensive. Using specific sections of the judgments for tasks such as summarization and descriptor assignment can improve the performance of these tasks.

As future work, we aim to expand the dataset to incorporate diverse structures of judgments. This will enable us to further specialize the model and enhance its performance.

Acknowledgments. This research was supported by Fundação para a Ciência e Tecnologia (FCT), through the INESC-ID multi-annual funding with reference DOI:10.54499/UIDB/50021/2020. This research is part of the IRIS project with reference PR07005. This project is a collaboration work involving the Portuguese Supreme Court of Justice and INESC-ID. This research was also supported by the Portuguese Recovery and Resilience Plan through project C645008882-00000055

References

1. Bhattacharya, P., Paul, S., Ghosh, K., Ghosh, S., Wyner, A.: Identification of rhetorical roles of sentences in indian legal judgments. In: Legal Knowledge and Information Systems - JURIX 2019: The Thirty-second Annual Conference, Madrid, Spain, December 11-13, 2019. Frontiers in Artificial Intelligence and Applications, vol. 322, pp. 3–12. IOS Press (2019)
2. de Carvalho, V.R., Cohen, W.W.: Learning to extract signature and reply lines from email. In: CEAS 2004 - First Conference on Email and Anti-Spam, July 30-31, 2004, Mountain View, California, USA (2004)
3. Chen, H., Hu, J., Sproat, R.: Integrating geometrical and linguistic analysis for email signature block parsing. ACM Trans. Inf. Syst. **17**(4), 343–366 (1999)
4. Chen, H., Cai, D., Dai, W., Dai, Z., Ding, Y.: Charge-based prison term prediction with deep gating network. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th International Joint Conf. on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019. pp. 6361–6366. ACL (2019)
5. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.P.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**, 2493–2537 (2011)
6. Dandapat, S., et al.: Part of specch tagging and chunking with maximum entropy model. In: Proc. of the IJCAI Workshop on Shallow Parsing for South Asian Languages. pp. 29–32 (2007)
7. Estival, D., Gaustad, T., Pham, S., Radford, W., Hutchinson, B.: Author profiling for english emails (12 2007)
8. Gael, J.V., Vlachos, A., Ghahramani, Z.: The infinite HMM for unsupervised pos tagging. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL. pp. 678–687. ACL (2009)

9. He, Z., Wang, Z., Wei, W., Feng, S., Mao, X., Jiang, S.: A survey on recent advances in sequence labeling from deep learning models. CoRR **abs/2011.06727** (2020)
10. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. CoRR **abs/1508.01991** (2015)
11. Jardim, B., Rei, R., Almeida, M.S.C.: Multilingual email zoning. In: Proc. of the 16th Conf. of the European Chapter of the ACL: Student Research Workshop, EACL 2021, Online, April 19-23, 2021. pp. 88–95. ACL (2021)
12. Jiang, W., Guan, Y., Wang, X.: Conditional random fields based label sequence and information feedback. In: Computational Intelligence, International Conference on Intelligent Computing, ICIC 2006, Kunming, China, August 16-19, 2006. Proceedings, Part II. Lecture Notes in Computer Science, vol. 4114, pp. 677–689. Springer (2006)
13. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proc. of the Eighteenth Intl. Conf. on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001. pp. 282–289. Morgan Kaufmann (2001)
14. Malik, V., Sanjay, R., Guha, S.K., Nigam, S.K., Hazarika, A., Bhattacharya, A., Modi, A.: Semantic segmentation of legal documents via rhetorical roles. CoRR **abs/2112.01836** (2021)
15. McCallum, A., Freitag, D., Pereira, F.C.N.: Maximum entropy markov models for information extraction and segmentation. In: Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, June 29 - July 2, 2000. pp. 591–598. Morgan Kaufmann (2000)
16. Melo, R., Santos, P.A., Dias, J.: A semantic search system for the supremo tribunal de justiça. In: Progress in Artificial Intelligence - 22nd EPIA Conference on Artificial Intelligence, EPIA 2023, Faial Island, Azores, September 5-8, 2023, Proc., Part II. Lecture Notes in Computer Science, vol. 14116, pp. 142–154. Springer (2023)
17. Nasr, A., Béchet, F., Volanschi, A.: Tagging with hidden markov models using ambiguous tags. In: COLING 2004, 20th International Conference on Computational Linguistics, Proc. of the Conf., 23-27 August 2004, Geneva, Switzerland (2004)
18. Qiao, M., Bian, W., Xu, R.Y.D., Tao, D.: Diversified hidden markov models for sequential labeling. IEEE Trans. Knowl. Data Eng. **27**(11), 2947–2960 (2015)
19. Ratnaparkhi, A.: A maximum entropy model for part-of-speech tagging. In: Conference on Empirical Methods in Natural Language Processing, EMNLP 1996, Philadelphia, PA, USA, May 17-18, 1996 (1996)
20. Savelka, J., Ashley, K.D.: Segmenting us court decisions into functional and issue specific parts. In: JURIX. pp. 111–120 (2018)
21. Turtle, H.R.: Text retrieval in the legal world. Artif. Intell. Law **3**(1-2), 5–54 (1995)
22. Wei, T., Qi, J., He, S., Sun, S.: Masked conditional random fields for sequence labeling. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021. pp. 2024–2035. ACL (2021)
23. Ye, H., Jiang, X., Luo, Z., Chao, W.: Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. In: Proc. of the 2018 Conference of the North American Chapter of the ACL: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers). pp. 1854–1864. ACL (2018)
24. Zhao, L., Qiu, X., Zhang, Q., Huang, X.: Sequence labeling with deep gated dual path CNN. IEEE ACM Trans. Audio Speech Lang. Process. **27**(12), 2326–2335 (2019)