Proposta recebida em Março 2025 e aceite para publicação em Maio 2025.

# Atribuição de Descritores a Acórdãos do Supremo Tribunal de Justiça Português com base em Representações Locais Esparsas

**SLEEC assignment of descriptors to judgments of the Supreme Court of Justice of Portugal**

Martim Zanatti
INESC-ID, Portugal
Instituto Superior Técnico, Universidade de Lisboa, Portugal

Ricardo Ribeiro
INESC-ID, Portugal
Iscte, Instituto Universitário de Lisboa, Portugal

H. Sofia Pinto
INESC-ID, Portugal
Instituto Superior Técnico, Universidade de Lisboa, Portugal

José Borbinha
INESC-ID, Portugal
Instituto Superior Técnico, Universidade de Lisboa, Portugal

**Resumo**

A Classificação Extrema Multi-etiqueta (XML) consiste na predição de múltiplas etiquetas para um determinado input, sendo um problema fundamental em domínios como categorização de texto, sistemas de recomendação e marcação de imagens. Esta tarefa apresenta desafios significativos para a aprendizagem automática e a recuperação de informação, especialmente devido ao crescimento exponencial de dados online e à consequente necessidade de algoritmos capazes de lidar com conjuntos de dados de grande escala e com um elevado número de etiquetas. Os métodos tradicionais de classificação são inadequados para esta tarefa devido ao vasto número de possíveis combinações de etiquetas e à dispersão das atribuições. Este artigo apresenta os resultados de um projeto realizado com o Supremo Tribunal de Justiça de Portugal, onde abordámos este problema utilizando *Sparse Local Embeddings for Extreme Multi-label Classification* (SLEEC), uma abordagem baseada em embeddings que demonstrou resultados promissores no domínio legal. O nosso objetivo foi associar descritores, que categorizam os acórdãos do tribunal Português, aos respetivos acórdãos. Este trabalho enfrentou diversos desafios, nos quais se incluem um elevado número de descritores, um conjunto de dados desbalanceado, a presença de muitas etiquetas raras (*tail labels*) e a extensão considerável dos documentos. Os resultados experimentais demonstram que a nossa abordagem alcançou uma variação de precisão/cobertura entre 0,57 e 0,68, indicando um desempenho promissor nesta tarefa complexa.

**Palavras chave**

descritores; documentos legais; classificação extrema multi-etiqueta; SLEEC

**Abstract**

Extreme Multi-label Classification (XML) involves predicting multiple labels for a given input, a fundamental problem in domains such as text categorization, recommendation systems, and image tagging. This task presents significant challenges for machine learning and information retrieval, particularly given the exponential growth of online data and the concomitant need for algorithms capable of handling large-scale datasets with numerous labels. Traditional classification methods are inadequate for this task due to the vast number of possible label combinations and the sparsity of label assignments. This paper reports the results of a project with the Supreme Court of Justice of Portugal ("Supremo Tribunal de Justiça Português") to address the problem using Sparse Local Embeddings for Extreme Multi-label Classification (SLEEC), an embedding-based approach that showed promising results in legal datasets. Our goal was to associate descriptors, which categorize court judgments, with the judgments themselves. This work tackled various challenges, including a large number of descriptors, an unbalanced dataset, numerous tail labels, and extensive document lengths. Our experimental results demonstrate that our approach achieved a precision/recall variation ranging between 0.57 and 0.68, indicating promising performance in this complex task.

DOI: 10.21814/lm.17.1.481

**Keywords**

descriptors; legal documents; extreme multi-label classification; SLEEC

## 1. Introdução

A Classificação Extrema Multi-etiqueta (Extreme Multi-label Classification) (XML) tornou-se um desafio crítico na aprendizagem automática e na recuperação de informação (Wei et al., 2022). A expansão exponencial de conteúdos textuais *online* exige algoritmos capazes de gerir de forma eficaz conjuntos de dados de grande escala com um elevado número de etiquetas. Esta tendência, aliada ao avanço generalizado da capacidade computacional, tem impulsionado progressos significativos neste domínio de investigação.

As tarefas de XML envolvem a predição de múltiplas etiquetas para uma determinada instância de *input*, sendo um problema fundamental em diversos domínios, como a categorização de texto, os sistemas de recomendação e a anotação de imagens. Por exemplo, na categorização de texto, os algoritmos XML desempenham um papel crucial na atribuição automática de etiquetas ou categorias relevantes a grandes volumes de dados textuais. No entanto, os métodos tradicionais de classificação são inadequados para tarefas XML devido ao número massivo de possíveis combinações de etiquetas e à dispersão das atribuições em conjuntos de dados do mundo real. Como resposta a esses desafios, surgiram vários algoritmos e técnicas específicas para XML, incluindo abordagens como a relevância binária (Khandagale et al., 2019; Niculescu-Mizil & Abbasnejad, 2017), métodos baseados em representações densas de textos, *embeddings* (Bhatia et al., 2015; Papanikolaou et al., 2016), abordagens baseadas em árvores (Agrawal et al., 2013; You et al., 2018), que operam sob a suposição de que existe uma estrutura hierárquica entre etiquetas, e arquiteturas de aprendizagem profunda adaptadas à classificação multi-etiqueta (Zhang et al., 2017; Wang et al., 2019).

Este trabalho resulta de uma colaboração entre o INESC-ID e o Supremo Tribunal de Justiça de Portugal[1] (STJ). O objetivo é associar descritores (termos que categorizam os acórdãos do Supremo Tribunal de Justiça) aos respetivos acórdãos. A contribuição deste trabalho resulta numa ferramenta que facilita a tarefa de atribuição de descritores aos acórdãos, tornando o fluxo de trabalho dos juízes mais eficiente e sistemático. A base deste trabalho é uma abordagem baseada em *embeddings* que obteve bons resultados no conjunto de dados jurídicos *EurLex*, com uma precisão no primeiro resultado (p@1) de 80,17, designada *Sparse Local Embeddings for Extreme Multi-label Classification* (SLEEC) (Bhatia et al., 2015). Outra contribuição relevante deste trabalho foi a utilização de partes específicas dos acórdãos. Os acórdãos do STJ podem ser divididos em secções, cada uma contendo diferentes aspetos do caso (Zanatti et al., 2024). Através da manipulação do espaço de *embeddings* dos acórdãos, procurámos compreender quais as partes dos acórdãos mais relevantes na atribuição dos descritores. A Secção 3 apresenta uma descrição detalhada do nosso conjunto de dados. Alguns desafios que importa mencionar foram, como é comum em tarefas similares, o elevado número de descritores, um conjunto de dados não balanceado, um número significativo de etiquetas raras e documentos extensos. Os melhores resultados, com uma variação de precisão/cobertura (ver Secção 5), situam-se entre 0,57 e 0,68.

O artigo está organizado da seguinte forma: após esta introdução, a Secção 2 apresenta uma visão geral do estado da arte relativamente à tarefa de XML, destacando os principais avanços e tendências da investigação nesta área; na Secção 3 descrevemos detalhadamente o conjunto de dados, evidenciando os desafios associados à atribuição de descritores aos acórdãos do STJ; a Secção 4 descreve o método proposto, explicando os algoritmos e técnicas utilizados. Na Secção 5 detalhamos a configuração experimental e apresentamos os resultados experimentais. Por fim, a Secção 6 conclui o artigo e discute possíveis direções para investigação futura.

## 2. Trabalho Relacionado

As tarefas convencionais de classificação multi-etiqueta têm como objetivo prever múltiplas etiquetas ou categorias, selecionadas a partir de um conjunto reduzido de etiquetas, para uma determinada instância. Formalmente, este problema pode ser definido do seguinte modo: dado um conjunto de instâncias $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, onde $x_i$ representa as características da $i$-ésima instância e $y_i$ é um vetor binário que indica a presença ou ausência de cada etiqueta, o objetivo é aprender um classificador $f(x)$ capaz de prever as etiquetas relevantes para novas instâncias não observadas $x$.

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

No entanto, em problemas XML, em vez de prever etiquetas para uma nova instância $x$ a partir de um conjunto reduzido de opções, é necessário selecionar entre milhares, ou até milhões de etiquetas, tornando a tarefa consideravelmente mais difícil.

### 2.1. Abordagens de Relevância Binária

Uma possível abordagem XML consiste em assumir que as etiquetas são independentes entre si e treinar classificadores binários individuais para cada etiqueta. O objetivo é minimizar independentemente a função de perda de cada classificador de etiquetas. No entanto, à medida que o número de etiquetas aumenta, a complexidade temporal e de memória cresce linearmente, resultando num elevado custo computacional. Para enfrentar estes desafios, diversas abordagens foram propostas, incluindo o uso de técnicas de computação paralela para distribuir o processamento entre múltiplos processadores ou núcleos (Babbar & Schölkopf, 2017, 2018); particionamento de etiquetas, que divide o conjunto de etiquetas em subconjuntos menores para reduzir a carga computacional (Khandagale et al., 2019); transformação de características, que consiste na redução da dimensionalidade ao transformar características num espaço de menor dimensão ou aplicar técnicas de seleção de características (Boutsidis et al., 2009); e aquecimento de classificadores, onde os classificadores são pré-treinados em subconjuntos de dados antes de serem ajustados no conjunto completo (Fang et al., 2019). Para reduzir o tempo de teste, trabalhos anteriores, como Niculescu-Mizil & Abbasnejad (2017), propuseram a préseleção de um subconjunto de etiquetas candidatas antes da aplicação do classificador multi-etiqueta base. Além disso, investigações recentes impuseram restrições de dispersão no modelo para reduzir o consumo de memória (Yen et al., 2016), remover parâmetros supérfluos para diminuir o tamanho do modelo (Babbar & Schölkopf, 2017, 2018; Khandagale et al., 2019) e aplicaram a técnica *Count-Min Sketch* para comprimir estruturas de dados de forma eficiente (Medini et al., 2019).

### 2.2. Abordagens Baseadas em Representações de Embeddings

Para lidar com o problema do grande número de etiquetas disponíveis, uma estratégia possível consiste em reduzir a sua dimensionalidade. Uma abordagem frequentemente utilizada para esse fim é a abordagem que consiste representações de texto baseadas em embeddings. Este tipo de abordagens procura diminuir o número efetivo de etiquetas projetando vetores de etiquetas num espaço de menor dimensão, assumindo que a matriz de etiquetas tem uma estrutura característica de menor valor. Diversas técnicas são utilizadas para abordar este problema, tais como filtros de Bloom (Cissé et al., 2013), Decomposição em Valores Singulares (SVD) (Tai & Lin, 2012) e etiquetas de referência (*landmark labels*) (Bi & Kwok, 2013). O método *Label Powerset* transforma um problema multi-etiqueta num problema multi-classe. Por exemplo, Papanikolaou et al. (2016) combina múltiplas etiquetas para criar novas etiquetas, convertendo assim o problema num de etiqueta única. Modelos de regressão são treinados para prever estes vetores projetados. Esta redução dos vetores de etiquetas proporciona simplicidade, uma base teórica sólida e a capacidade de lidar com correlações entre etiquetas. No entanto, os métodos que consistem em representações de texto baseadas em embeddings são lentos no treino e na predição, e a suposição de que a matriz de etiquetas de treino tem uma característica de menor valor nem sempre se verifica em cenários reais.

Para superar essas limitações, SLEEC (Bhatia et al., 2015) expande os métodos que consistem em representações de texto baseadas em embeddings ao aprender representações que capturam correlações não-lineares entre etiquetas, preservando, assim, apenas as distâncias entre os vetores de etiquetas mais próximos. Esta abordagem difere dos métodos tradicionais de projeção linear em sub-espaços ao focar-se na preservação das distâncias locais em vez de uma aproximação global de característica de menor valor. O SLEEC introduz uma nova formulação para a aprendizagem de representações de texto baseadas em embeddings que garante a preservação dos vizinhos mais próximos no espaço de etiquetas, proporcionando um processo de treino eficiente e, muitas vezes, mais rápido do que os métodos tradicionais que usam embeddings como abordagem. Para lidar com conjuntos de dados de grande escala e espaços de etiquetas de alta dimensão, o SLEEC utiliza técnicas de clustering para dividir os dados em subconjuntos mais pequenos e manejáveis. Ao aprender representações de texto baseadas em embeddings e classificadores localmente dentro de cada cluster, o SLEEC melhora significativamente a precisão das predições e a escalabilidade do modelo. Para lidar com possíveis instabilidades no clustering, o SLEEC emprega *ensemble learning*, onde cada modelo é gerado por uma partição aleatória diferente dos

dados, ajudando assim a estabilizar o processo. Durante a predição, o SLEEC utiliza um classificador k-nearest neighbor (kNN) no espaço de embeddings, tirando partido da preservação dos vizinhos mais próximos durante o treino. Para novos pontos de dados, o vetor de etiquetas previsto é obtido através da agregação das etiquetas dos seus vizinhos mais próximos no espaço de embeddings, em vez de considerar todos os pontos de treino.

Inspirando-se no SLEEC, AnnexML (Tagami, 2017) introduziu uma metodologia de embeddings baseada em grafos, onde é construído um grafo de kNN com base nos vetores de etiquetas, procurando replicar fielmente a estrutura do grafo dentro do espaço de embeddings. A predição é realizada de forma eficiente através do uso de técnicas de busca aproximada de vizinhos mais próximos (*Approximate Nearest Neighbor Search*, ANNS) (Dong et al., 2011), melhorando significativamente a escalabilidade.

As redes neuronais profundas têm recebido uma atenção significativa nos últimos anos devido à sua capacidade expressiva. A abordagem DeepXML (Zhang et al., 2017) aprofunda a análise do espaço de etiquetas ao construir um grafo explícito de etiquetas e ao aprender representações de texto baseados em embeddings não-lineares para características e etiquetas. Esta abordagem visa capturar relações complexas tanto no espaço das características como no das etiquetas, melhorando assim a precisão das predições. Por outro lado, o Rank-AE (Wang et al., 2019) propõe um *AutoEncoder* baseado em ranking para explorar dependências entre etiquetas e relações entre características e etiquetas. Este método projeta etiquetas e características num espaço de embeddings comum, facilitando uma compreensão mais profunda dos dados subjacentes.

## 2.3. Abordagens Baseadas em Árvores

As abordagens baseadas em árvores operam sob a suposição de que existe uma estrutura hierárquica entre as etiquetas. Uma vantagem notável destas metodologias é o seu tempo de inferência consideravelmente reduzido, que geralmente escala de forma logarítmica em relação ao número de etiquetas. Esta abordagem pode ser dividida em dois tipos principais: árvores de instâncias e árvores de etiquetas.

Nas árvores de instâncias, cada nó contém um subconjunto dos exemplos de treino, posteriormente distribuídos para os nós filhos com base em determinados critérios. A ideia subjacente é a seguinte: em cada nó apenas um subconjunto de etiquetas – especificamente, as etiquetas que estão atribuídas aos pontos de treino desse nó – estão presentes. Isto permite um processamento mais focado e eficiente dos dados de treino em cada nó, uma vez que apenas as etiquetas relevantes precisam de ser consideradas.

Por outro lado, as árvores de etiquetas organizam as etiquetas hierarquicamente, onde cada nó contém somente um conjunto de etiquetas distribuído entre os seus nós filhos. A estrutura da árvore é determinada através de *clustering* recursivo das etiquetas até que sejam satisfeitas condições terminais. A clusterização de etiquetas visa dividir o conjunto de etiquetas em subconjuntos disjuntos com base na sua similaridade, uma tarefa frequentemente realizada por algoritmos como o *K-means*. Durante a predição, uma nova instância percorre uma árvore de instâncias até um nó folha e uma árvore de etiquetas até vários nós que contêm etiquetas. Na árvore de etiquetas, esses vários nós representam as etiquetas de saída consideradas para a predição. No entanto, nas árvores de instâncias, a predição é feita por um classificador treinado sobre os exemplos associados ao nó folha. Este mecanismo garante que a predição seja baseada num subconjunto específico de exemplos de treino, permitindo previsões mais direcionadas e precisas.

### *2.3.1. Árvores de Instâncias*

A abordagem *Multi-label Classification Through Random Forest* (MLRF) (Agrawal et al., 2013) aborda problemas com grandes conjuntos de etiquetas ao dividir o espaço de características do nó pai entre os seus nós filhos, reduzindo assim o número de pontos de dados de treino e etiquetas em cada nó filho. Por outro lado, o FastXML (Prabhu & Varma, 2014) aprende a hierarquia otimizando a perda baseada no ranking *normalized Discounted Cumulative Gain* (nDCG). O *nDCG* apresenta duas vantagens principais: pode ser otimizado em todas as etiquetas no nó atual, garantindo que a otimização local não seja míope, e seja sensível à ordenação e relevância, assegurando que as etiquetas positivas relevantes sejam previstas com classificações tão elevadas quanto possível.

Para lidar com as etiquetas raras, o PfastreXML (Jain et al., 2016), baseado no FastXML, substitui a perda *nDCG* por uma variante corrigida por propensão. Esta modificação atribui recompensas mais altas para previsões corretas de etiquetas raras, melhorando assim o desempenho do modelo nesses casos mais desafiantes.

### *2.3.2. Árvores de Etiquetas*

O Homer (*Hierarchy Of Multilabel classifiERs)* (Tsoumakas et al., 2008) introduz um classificador hierárquico multi-etiqueta utilizando o algoritmo *k-means* balanceado para particionar recursivamente o conjunto de etiquetas de forma equilibrada. Cada nó está associado a um classificador multi-etiqueta treinado para prever as etiquetas dos exemplos. O AttentionXML (You et al., 2018) integra representações Glove (Pennington et al., 2014) e um mecanismo de atenção no XML, permitindo a aprendizagem de informação semântica a partir de dados textuais brutos utilizando redes *Long Short-Term Memory* (LSTM). Além disso, utiliza múltiplas árvores probabilísticas para melhorar a eficiência do treino.

Os métodos baseados em árvores enfrentam desafios na precisão da predição devido ao efeito em cascata, onde erros ocorridos nos níveis superiores da árvore não podem ser corrigidos nos níveis inferiores. Esta limitação surge porque as decisões tomadas nos níveis superiores influenciam as subsequentes nos níveis inferiores, potencialmente amplificando imprecisões ao longo do processo de predição.

## 3. Conjunto de Dados

Como indicado na Secção 1, este trabalho foi realizado no âmbito do projeto *IRIS*, uma colaboração entre o INESC-ID e o Supremo Tribunal de Justiça (STJ). Conforme descrito na Secção 1, o objetivo foi associar automaticamente descritores aos acórdãos do STJ.

Os descritores são palavras-chave associadas a um acórdão que sintetizam as suas principais características. Estes descritores desempenham um papel fundamental na categorização dos acórdãos e são de extrema importância na investigação jurídica e na recuperação de informação, facilitando a pesquisa e a compreensão eficiente dos documentos jurídicos. O STJ mantém uma lista oficial de descritores, atualizada anualmente, para classificar os acórdãos. Utilizámos a lista oficial de 2023, que contém um total de 4815 descritores.

Os acórdãos do STJ são categorizados em quatro áreas distintas, cada uma especializada em diferentes tipos de casos e questões jurídicas:

- **Cível**: Abrange casos de natureza civil, incluindo disputas contratuais.
- **Criminal**: Engloba processos de natureza penal, como homicídios.
- **Social**: Refere-se a casos na área social, incluindo questões relacionadas com pensões.
- **Contencioso**: Trata de litígios entre cidadãos e entidades públicas ou entre diferentes entidades públicas.

Estas áreas estão estruturadas em secções, onde o número de cada secção representa o momento temporal relativo da sua criação, refletindo a evolução do número de processos apresentados ao STJ. Atualmente, a área **Cível** possui quatro secções (**Primeira, Segunda, Sexta e Sétima**), a área **Criminal** tem duas (**Terceira e Quinta**) e a área **Social** conta com uma única secção (**Quarta**). A área **Contencioso** é especial, uma vez que trata apenas de casos raros e específicos. Esta separação tem como objetivo estabelecer secções especializadas para lidar com diferentes tipos de processos ou matérias jurídicas específicas. Tal organização melhora a eficiência e a especialização na administração da justiça, permitindo que juízes e magistrados se concentrem em áreas específicas do direito conforme a sua experiência. A segmentação de cada área em secções especializadas facilita o tratamento célere e preciso dos processos, possibilitando aos juízes um maior domínio sobre matérias jurídicas específicas. Além disso, esta segmentação garante uma distribuição mais equitativa dos processos entre os diferentes juízes e magistrados. Os casos (e os respetivos acórdãos) têm bem definido quanto à área e secção a que pertencem, de modo a serem atribuídos a um conjunto de juízes especializados na respetiva área e secção.

O conjunto de dados utilizado[2] contém um total de 26870 acórdãos e os respetivos descritores, provenientes das oito origens mencionadas anteriormente (as sete secções mais a área de Contencioso).

Os documentos jurídicos tendem a ser mais extensos do que documentos de outros domínios (Turtle, 1995; Kanapala et al., 2019). No nosso caso, isto é particularmente evidente, uma vez que estamos a trabalhar com acórdãos da terceira instância (Supremo Tribunal), que frequentemente incluem o conteúdo do processo das instâncias inferiores (primeira instância ou Tribunal de Comarca, e segunda instância ou Tribunal da Relação). O tamanho médio dos acórdãos no nosso conjunto de dados é de 7500 palavras.

Cada documento possui um ou mais descritores. No entanto, a utilização dos descritores não seguiu um padrão conforme os critérios da lista

[2] https://huggingface.co/datasets/MartimZanatti/Descriptors_STJ

oficial, uma vez que o sistema onde são registados não impõe essa padronização. O mesmo descritor pode ser expresso de várias formas, incluindo diferenças na capitalização, variações no uso de *stopwords* como *de*, *da*, *do*, e variações entre formas no singular e plural. Realizámos, portanto, a normalização dos descritores convertendo-os para letras minúsculas, eliminando *stopwords* e aplicando lematização. Após este processo de normalização, efetuámos uma verificação manual dos descritores para resolver casos mais complexos. No final, os descritores que não constavam na lista oficial foram descartados. A Tabela 1 e a Tabela 2 apresentam o número de acórdãos e descritores por área e secção, respetivamente.

| Área | No. de Acordãos | No. de Descritores |
|---|---|---|
| *Cível* | 14978 | 3313 |
| *Criminal* | 7726 | 1771 |
| *Social* | 3539 | 1404 |
| *Contencioso* | 279 | 504 |

**Tabela 1:** Número de acórdãos e descritores por área.

| Secção | No. de Acordãos | No. de Descritores |
|---|---|---|
| *Primeira Secção (Cível)* | 3997 | 2496 |
| *Segunda Secção (Cível)* | 4040 | 2485 |
| *Terceira Secção (Criminal)* | 4750 | 1560 |
| *Quarta Secção* (Social) | 3539 | 1404 |
| *Quinta Secção (Criminal)* | 2976 | 1283 |
| *Sexta Secção (Cível)* | 3004 | 2262 |
| *Sétima Secção (Cível)* | 3937 | 2500 |
| *Contencioso* | 279 | 504 |

**Tabela 2:** Número de acórdãos e descritores por secção.

Para além do elevado número de descritores, observa-se um claro desbalanceamento na sua distribuição. Um exemplo desta tendência é apresentado na Figura 1 para a área Cível. Todas as áreas e secções exibem um padrão semelhante (as restantes áreas não são apresentadas aqui devido a limitações de espaço). Destaca-se que uma proporção significativa de descritores (neste caso, 496 na área Cível) é utilizada apenas uma vez nos acórdãos da respetiva área. Além disso, a frequência de descritores que aparecem múltiplas vezes nos acórdãos decresce exponencialmente em todas as áreas e secções. A complexidade da tarefa advém da vasta gama de descritores possíveis e da sua distribuição altamente desbalanceada. Juntando ao pré-processamento explicado no parágrafo anterior, também excluímos do conjunto de dados, para cada área e secção, os descritores que surgem somente uma ou duas vezes nas respetivas áreas e secções, pois torna-se impraticável incluí-los nos conjuntos de treino, teste e desenvolvimento. Nas Tabelas 3 e 4 apresentam-se os números de acórdãos e descritores por área e por secção, respetivamente, após a remoção dos descritores que surgem apenas uma ou duas vezes. Importa referir que a remoção de acórdãos se deve ao facto de que, caso os respetivos descritores ocorram apenas uma ou duas vezes e sejam removidos, esses acórdãos ficam sem qualquer descritor associado — motivo pelo qual são também excluídos do conjunto de dados.

| Área | No. de Acordãos | No. de Descritores |
|---|---|---|
| *Cível* | 14966 | 2503 |
| *Criminal* | 7713 | 1105 |
| *Social* | 3504 | 725 |
| *Contencioso* | 279 | 439 |

**Tabela 3:** Número de ácordãos e descritores por área, após a remoção de descritores que aparecem somente uma ou duas vezes.

| Secção | No. de Acordãos | No. de Descritores |
|---|---|---|
| *Primeira Secção (Cível)* | 3973 | 1474 |
| *Segunda Secção (Cível)* | 4022 | 1433 |
| *Terceira Secção (Criminal)* | 4738 | 931 |
| *Quarta Secção (Social)* | 3504 | 725 |
| *Quinta Secção (Criminal)* | 2968 | 725 |
| *Sexta Secção (Cível)* | 2986 | 1225 |
| *Sétima Secção (Cível)* | 3927 | 1524 |
| *Contencioso* | 279 | 439 |

**Tabela 4:** Número de ácordãos e descritores por secção, após a remoção de descritores que aparecem somente uma ou duas vezes.

A Tabela 5 e a Tabela 6 ilustram as matrizes de confusão para as secções e áreas, respetivamente. Ao treinar um modelo para cada área e cada secção, conseguimos reduzir o número de descritores. Prevê-se que as secções dentro da mesma área apresentem uma maior interseção de descritores, dada a semelhança nos temas jurídicos abordados.

Para reduzir o volume de descritores, explorámos várias abordagens que são simultaneamente práticas e teoricamente fundamentadas. De facto, é expectável que cada área utilize descritores distintos, uma vez que se foca em diferentes tipos de casos e questões jurídicas. Por exemplo, descritores como *"cibercrime"* poderão

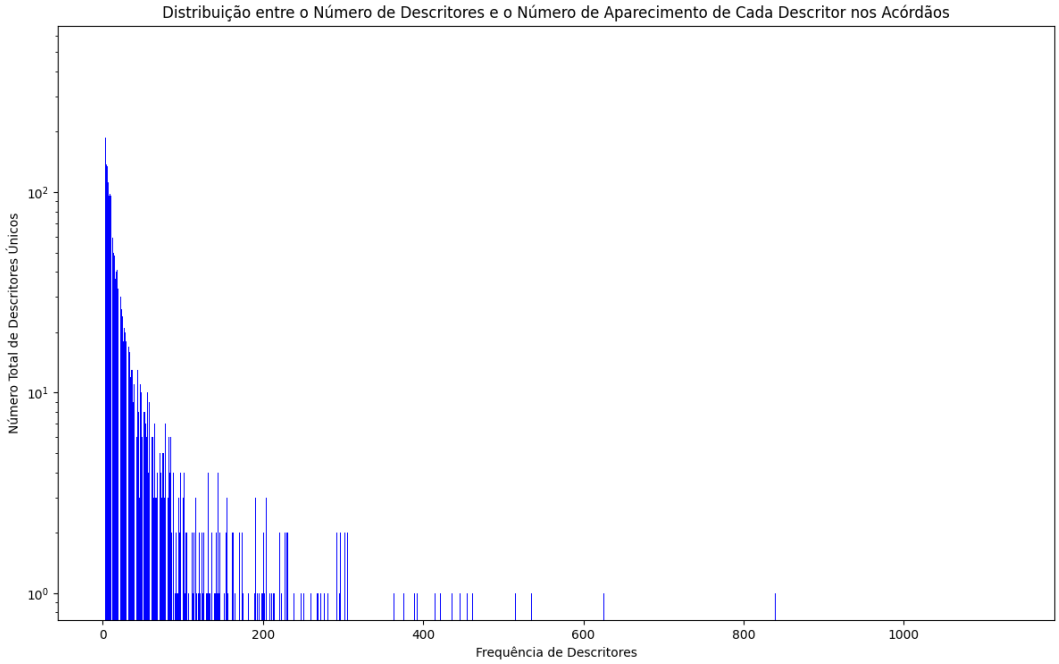


**Figura 1:** Distribuição entre o número de descritores e o número de aparecimento de cada descritor nos ácordãos da área Cível.

| Secção | Primeira Secção | Segunda Secção | Terceira Secção | Quarta Secção | Quinta Secção | Sexta Secção | Sétima Secção | Contencioso |
|---|---|---|---|---|---|---|---|---|
| Primeira Secção | 2496 | 2031 | 963 | 964 | 759 | 1872 | 2047 | 364 |
| Segunda Secção | 2031 | 2485 | 942 | 943 | 730 | 1887 | 2043 | 352 |
| Terceira Secção | 963 | 942 | 1560 | 629 | 1072 | 876 | 943 | 333 |
| Quarta Secção | 963 | 943 | 629 | 1404 | 500 | 888 | 960 | 317 |
| Quinta Secção | 759 | 730 | 1072 | 500 | 1283 | 697 | 736 | 299 |
| Sexta Secção | 1872 | 1887 | 876 | 888 | 697 | 2262 | 1900 | 340 |
| Sétima Secção | 2047 | 2043 | 943 | 960 | 736 | 1900 | 2500 | 356 |
| Contencioso | 364 | 352 | 333 | 317 | 299 | 340 | 356 | 504 |

**Tabela 5:** Intersecção de descritores entre secções.

| Área | Cível | Criminal | Social | Contencioso |
|---|---|---|---|---|
| Cível | 3313 | 1321 | 1126 | 421 |
| Criminal | 1321 | 1771 | 687 | 364 |
| Social | 1126 | 687 | 1404 | 317 |
| Contencioso | 421 | 364 | 317 | 504 |

**Tabela 6:** Intersecção de descritores entre áreas.

surgir exclusivamente na área Criminal e nas suas respetivas secções, refletindo a natureza específica dos casos tratados. De igual modo, a separação em secções visa abordar diferentes tipos de processos e matérias jurídicas, contribuindo para a divergência dos descritores utilizados em cada secção. Treinar um modelo específico para cada área e cada secção surge, assim, como uma abordagem promissora, tanto do ponto de vista teórico como prático. Esta estratégia reduz naturalmente o número de descritores associados a cada modelo, permitindo uma melhor adaptação ao domínio jurídico específico de cada área. Além disso, esta abordagem ajuda a mitigar a complexidade inerente à resolução do problema de XML. Ao treinar modelos separados para cada área e secção, o espaço da classificação é restringido, simplificando assim a tarefa global e melhorando o desempenho dos modelos.

## 4. Método

O método proposto assenta nos embeddings gerados pelo SLEEC, conforme referido anterior-

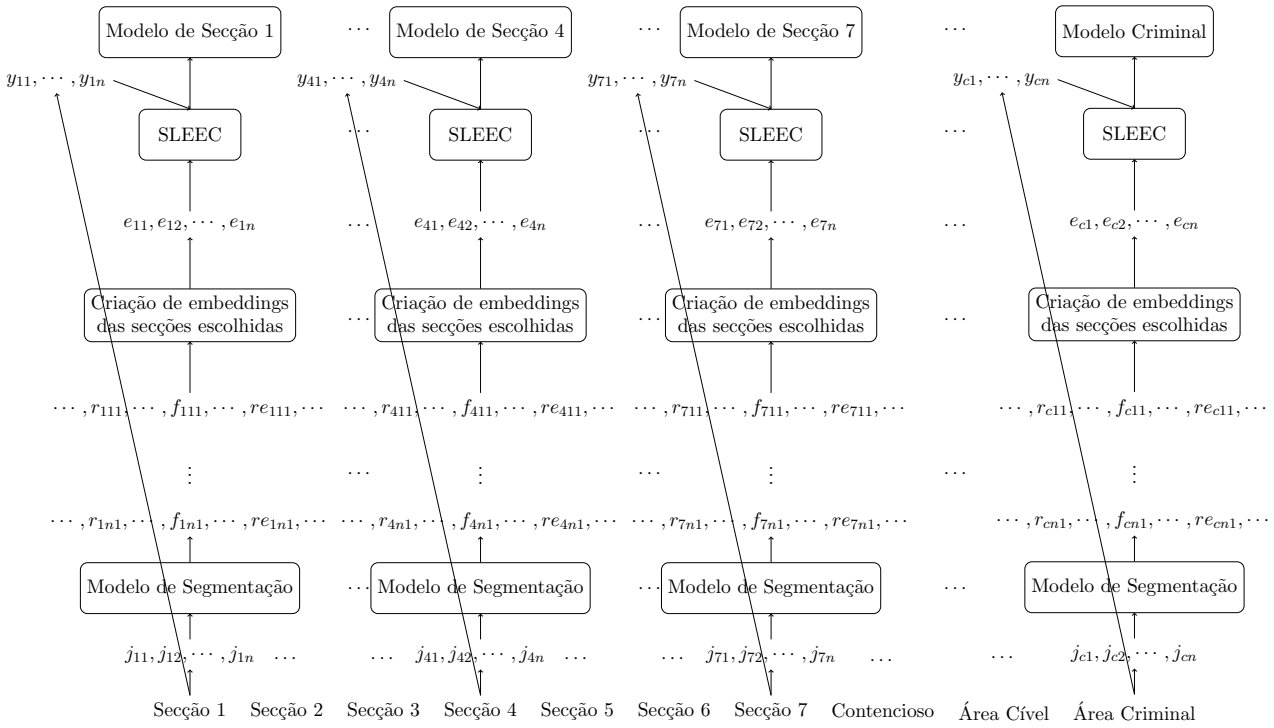


**Figura 2:** Treinámos um modelo SLEEC (Bhatia et al., 2015) para cada secção/área do STJ, onde $y$ representa as etiquetas de cada secção/área e $e$ representa os embeddings. Neste contexto, $j$ designa o *acórdão*, $r$ refere-se à secção Relatório, $f$ indica a secção Fundamentação de Facto e $re$ representa a Fundamentação de Direito, encapsulando as secções inerentes aos acórdãos do STJ (Zanatti et al., 2024). Uma descrição detalhada do funcionamento da nossa arquitetura é apresentada na Secção 4.1

.

mente. Optámos por este modelo porque, como mencionado na Secção 1, apresentou resultados promissores no domínio jurídico, alcançando um valor de p@1 de 80.17 no conjunto de dados *EurLex*. Além disso, o seu custo computacional é inferior ao de abordagens mais recentes baseadas em transformers, o que é particularmente relevante em contextos com recursos limitados. Uma dessas abordagens, o *AttentionXML* (You et al., 2018), atingiu um p@1 de 87.12 no mesmo conjunto de dados, mas, considerando o equilíbrio entre desempenho e custo computacional, optámos pelo SLEEC. Acresce que o SLEEC tem demonstrado bom desempenho em conjuntos de dados desbalanceados e com um elevado número de *tail labels*, características frequentemente observadas em corpora legais.

Treinámos um modelo SLEEC para cada secção e área, utilizando os respetivos acórdãos e descritores da secção e área, conforme ilustrado na Figura 2. O processo para cada secção e área decorreu da seguinte forma: cada acórdão foi representado como texto em formato HTML. Procedemos à divisão dos acórdãos em parágrafos utilizando o pacote Python *beautifulsoup4*.[3] O modelo de Segmentação de Acórdãos categorizou cada parágrafo de acordo com a sua secção relevante no acórdão (Zanatti et al., 2024). Alimentar o SLEEC apenas com secções do acórdão específicas permitiu reduzir o tamanho dos mesmos, que tendem a ser extensos, concentrando-se nas partes essenciais para a previsão dos descritores. Utilizámos o modelo *stjiris/bert-large-portuguese-cased-legal-mlm-nli-sts-v1*[4] (Melo et al., 2023) para codificar os parágrafos em embeddings. Este modelo é uma adaptação jurídica do BERTimbau large (Souza et al., 2020), representando cada parágrafo num espaço vetorial denso de 1024 dimensões. Treinado com a técnica Masked Language Model (MLM), utilizou uma taxa de aprendizagem de $1e-5$ e um *batch size* de 16. O conjunto de treino incluiu sentenças jurídicas extraídas de mais de 30000 documentos, abrangendo um total de 15000 passos de treino. Para representar um acórdão, somámos e normalizámos os *embeddings* dos respetivos parágrafos. Os embeddings resultantes, juntamente com as etiquetas correspondentes a cada acórdão, foram então introduzidas no modelo SLEEC.

[3] beautifulsoup4=4.11.1

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

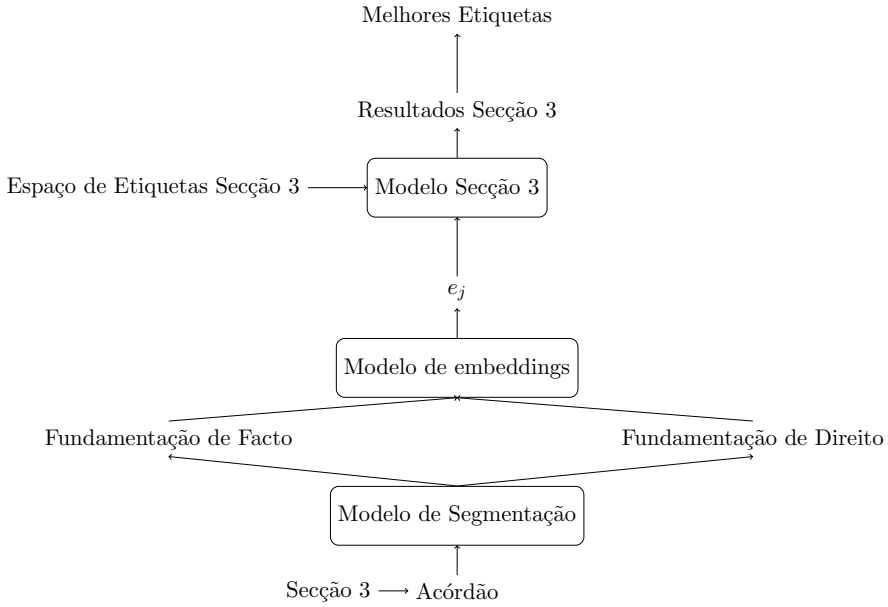


**Figura 3:** Caminho (1) de previsão de etiquetas para um acórdão da Secção 3 da Área Criminal: nesta ilustração, utilizámos as secções Fundamentação de Facto e Fundamentação de Direito como espaço de embeddings. Para prever as etiquetas do respetivo acórdão, aplicámos o modelo treinado para a Secção 3 da Área Criminal (englobando todos os descritores da respetiva área). As etiquetas com as melhores pontuações são retornadas. O número de etiquetas devolvidas pode ser ajustado conforme necessário.

### 4.1. Modelo de Segmentação de Acórdãos

Os acórdãos do STJ podem ser divididos em secções, cada uma desempenhando um papel distinto no caso. Estas secções incluem: **Cabeçalho, Relatório, Delimitação, Fundamentação de Facto, Fundamentação de Direito, Decisão, Assinatura, Declaração** e **Footnotes**. As funções específicas de cada secção são explicadas em detalhe em Zanatti et al. (2024). Algumas secções são separadas por um título; no entanto, a nomenclatura destes títulos pode variar consoante o juiz que redigiu o acórdão, e alguns acórdãos não possuem qualquer título. Por estas razões, não é viável segmentar as secções apenas com base no significado dos títulos. Para dividir automaticamente os parágrafos em secções, associando cada parágrafo à respetiva secção, foi treinado um modelo BI-LSTM-CRF. Os detalhes do modelo e do conjunto de dados utilizado são apresentados por Zanatti et al. (2024).

### 4.2. SLEEC

Tal como discutido na Secção 4, adotámos a abordagem SLEEC (Bhatia et al., 2015) para resolver o problema de XML. Ao contrário da metodologia original, que utiliza *clustering* com K-means[5] para particionar o conjunto de dados em subconjuntos menores, dividimos o conjunto de dados por secções e áreas, treinando um modelo distinto para cada uma. Mantivemos os hiperparâmetros do SLEEC fixos para todos os modelos treinados, definindo a dimensão das representações de embeddings (*embedding dimension*) em 560, o parâmetro $k$ no kNN em 15 e o número de vizinhos considerados durante a *Singular Value Projection (SVP)*, uma técnica de preenchimento de matrizes e aproximação de característica de menor valor (Jain et al., 2010), em 30.

Tomando como exemplo a Área Criminal, para prever as etiquetas de um acórdão desta área, podemos seguir três caminhos diferentes: (1) Verificar a que secção pertence – se à Secção 3 ou à Secção 5, que está explícito no próprio acórdão – e usar o modelo treinado somente para a respetiva secção, como é possível observar no exemplo da Figura 3. (2) Usar o exemplo da Figura 4, que implica usar o modelo treinado para a Área Criminal – que foi treinado tanto com os dados referentes à Secção 3 como os dados referentes à Secção 5 – para prever as etiquetas. (3) Usar o processo exemplificado na Figura 5, e tirar proveito do modelo treinado somente com

[5] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

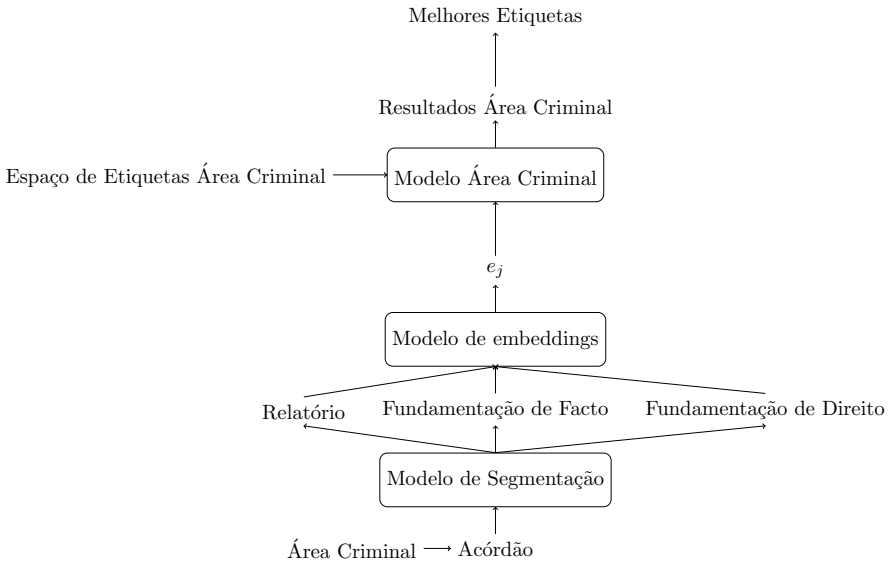


**Figura 4:** Caminho (2) de previsão de etiquetas para um acórdão da Área Criminal: nesta ilustração, utilizámos as secções Relatório, Fundamentação de Facto e Fundamentação de Direito como espaço de embeddings. Para prever as etiquetas do respetivo acórdão, aplicámos o modelo treinado para a Área Criminal (englobando todos os descritores da respetiva área). As etiquetas com as melhores pontuações são retornadas. O número de etiquetas devolvidas pode ser ajustado conforme necessário.

o conjunto de dados da Secção 3 e do modelo treinado somente com o conjunto de dados da Secção 5. Para modificar o espaço de embeddings dos modelos SLEEC, um aspeto crucial desta abordagem, testámos diferentes secções dos acórdãos. Ao segmentar os acórdãos em secções através do modelo de segmentação, utilizámos estas secções como conjunto de dados para treinar o modelo SLEEC, alterando assim o espaço de embeddings. Curiosamente, os resultados obtidos para secções individuais e áreas individuais foram inferiores aos do exemplo apresentado na Figura 5. Neste cenário, ao considerar um acórdão da Área Criminal, o melhor desempenho foi alcançado ao utilizar os modelos das secções e somar as pontuações das etiquetas. Uma análise mais detalhada é apresentada na Secção 5.

## 5. Resultados e Discussão

Nesta secção, apresentamos os resultados (de uma única experimentação) do nosso sistema utilizando várias configurações: treinámos um modelo para cada secção e para cada área, explorando diferentes configurações ao manipular o espaço de *embeddings*. As configurações consideradas são as seguintes:

- **r**: Utilização da secção Relatório como espaço de embeddings;
- **r todos os modelos**: Utilização da secção Relatório como espaço de embeddings e de todos os modelos das secções da respetiva área (como descrito na **Secção 5**);
- **l**: Utilização da secção de Fundamentação de Direito como espaço de embeddings;
- **l todos os modelos**: Utilização da secção de Fundamentação de Direito como espaço de embeddings e de todos os modelos das secções da respetiva área;
- **d**: Utilização da secção Decisão como espaço de embeddings;
- **d todos os modelos**: Utilização da secção Decisão como espaço de embeddings e de todos os modelos das secções da respetiva área;
- **f-l**: Utilização das secções de Fundamentação de Facto e Fundamentação de Direito como espaço de embeddings;
- **f-l todos os modelos**: Utilização das secções de Fundamentação de Facto e Fundamentação de Direito como espaço de embeddings e de todos os modelos das secções da respetiva área;
- **r-f-l**: Utilização das secções Relatório, Fundamentação de Facto e Fundamentação de Direito como espaço de embeddings;
- **r-f-l todos os modelos**: Utilização das secções Relatório, Fundamentação de Facto e Fundamentação de Direito como espaço de embeddings e de todos os modelos das secções da respetiva área;
- **r-l**: Utilização das secções Relatório e Fundamentação de Direito como espaço de embeddings;

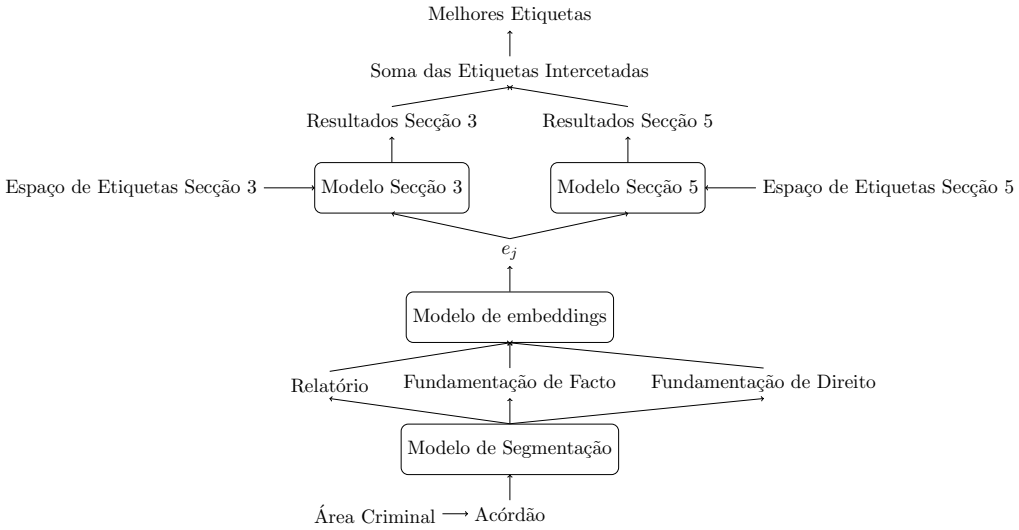


**Figura 5:** Caminho (3) de previsão de etiquetas para um acórdão da Área Criminal: nesta ilustração, utilizámos as secções Relatório, Fundamentação de Facto e Fundamentação de Direito como espaço de embeddings. Para prever as etiquetas do respetivo acórdão, aplicámos os modelos treinados para as secções específicas da Área Criminal. Somámos as pontuações das etiquetas que aparecem em mais de um modelo de secção. As etiquetas com as melhores pontuações são retornadas. O número de etiquetas devolvidas pode ser ajustado conforme necessário.

- **r-l todos os modelos**: Utilização das secções Relatório e Fundamentação de Direito como espaço de embeddings e de todos os modelos das secções da respetiva área;
- **r-l-d**: Utilização das secções Relatório, Fundamentação de Direito e Decisão como espaço de embeddings;
- **r-l-d todos os modelos**: Utilização das secções Relatório, Fundamentação de Direito e decisão como espaço de embeddings e de todos os modelos das secções da respetiva área;
- **todas as secções**: Utilização de todas as secções do acórdão como espaço de embeddings;
- **todas as secções/todos os modelos**: Utilização de todas as secções como espaço de embeddings e de todos os modelos das secções da respetiva área.

Estas configurações permitem avaliar a importância de cada secção do acórdão na atribuição de descritores ao respetivo acórdão. Calculámos uma variação das fórmulas de precisão e cobertura, com o sistema a retornar entre 1, 2, 3, 4, 5, 10, 15 e 20 descritores. Para avaliar com precisão o desempenho do nosso sistema, utilizamos uma variação da métrica precisão/cobertura (p/r), que segue a fórmula clássica da precisão quando o número de descritores retornados é inferior ao número de descritores associados ao acórdão. Por outro lado, quando o número de descritores retornados é superior ao número de descritores do acórdão, a métrica funciona como a fórmula clássica de cobertura. A variação da métrica p/r é definida da seguinte forma:

$$\text{p/r variation} = \frac{TP}{TP + \min(FP, len(\text{true descriptors}) - TP)} \quad (1)$$

Considere, por exemplo, um acórdão que tem cinco descritores associados. Se o nosso sistema retornar um descritor e este for um dos cinco descritores corretos, temos $TP = 1$ e descritores verdadeiros $= 5$. Assim, a variação da métrica p/r seria: $\frac{1}{1+min(0,4)} = 1$. Isto equivale à fórmula clássica da precisão: $\frac{TP}{TP+FP} = \frac{1}{1+0} = 1$. No entanto, a fórmula clássica de cobertura seria: $\frac{TP}{TP+FN} = \frac{1}{1+4} = \frac{1}{5}$. Ou seja, não é possível que o sistema retorne corretamente os cinco descritores quando apenas devolve um. Se o sistema retornar 20 descritores e 3 deles forem 3 dos 5 descritores corretos associados ao acórdão, então temos $TP = 3$ e descritores verdadeiros $= 5$. A fórmula da variação p/r seria: $\frac{3}{3+min(17,2)} = \frac{3}{5}$. Isto equivale à fórmula clássica de cobertura: $\frac{TP}{TP+FN} = \frac{3}{3+2} = \frac{3}{5}$. No entanto, a precisão clássica seria: $\frac{TP}{TP+FP} = \frac{3}{3+17} = \frac{3}{20}$. Ou seja, quando o sistema retorna mais descritores do que os descritores verdadeiros, a precisão é negativamente influenciada.

A Tabela 7 apresenta os resultados para as secções e áreas dentro do domínio Cível. Os melhores resultados para cada secção/área estão destacados a **negrito**. A partir destes resultados,

duas observações tornam-se evidentes: os melhores desempenhos são obtidos quando a **secção de Fundamentação de Direito** é considerada em combinação com a abordagem descrita na Figura 5. Mais especificamente, as configurações que alcançaram os melhores resultados foram **l, r-f-l, r-l** e **todas as secções**.

Surpreendentemente, a configuração **r-l-d** apresentou um desempenho inferior à configuração **r-l**. Isto pode ser atribuído ao facto de que a secção de Decisão é geralmente muito breve, contendo apenas a decisão final, e sendo escrita de forma semelhante em diferentes acórdãos. Quando incluída no espaço de embeddings, esta secção pode agrupar acórdãos que, na realidade, não partilham os mesmos descritores. Em outras palavras, a secção de Decisão não contribui para a seleção de descritores e pode até impactar negativamente a sua atribuição. Utilizar apenas a secção de Decisão para atribuir descritores resultou num desempenho significativamente inferior em comparação com outras configurações.

De forma semelhante, a configuração **r** também obteve alguns dos piores resultados em todas as secções/áreas dentro do domínio Cível. A secção Relatório introduz o conteúdo do acórdão, descreve as pretensões das partes envolvidas e as decisões a serem consideradas. Seria expectável que esta secção ajudasse na associação de descritores aos acórdãos, uma vez que os detalhes do caso são apresentados nesta parte do texto. Por exemplo, num caso relacionado com um acidente de viação, que corresponde a um descritor, a informação sobre o caso aparece geralmente na secção Relatório. No entanto, utilizar apenas a secção Relatório como espaço de embeddings revelou uma deficiência na associação de descritores quando comparado com outras configurações. Com exceção da Secção 1, onde a diferença é insignificante, a diferença entre a configuração **r** e a configuração com melhor desempenho varia entre **0,08** e **0,17**.

Curiosamente, quando a secção Relatório é combinada com a secção de Fundamentação de Direito e a secção de Fundamentação de Facto, as configurações **r-f-l** e **r-l** obtiveram resultados comparáveis aos melhores desempenhos. Os resultados evidenciam que a secção de Fundamentação de Direito, que apresenta a fundamentação para a decisão final do acórdão, é a mais relevante na atribuição de descritores. No entanto, em alguns casos, as configurações **r-f-l** e **r-l** superaram o desempenho da configuração que utiliza apenas a secção de Fundamentação de Direito como espaço de embeddings. Isto sugere que as secções Relatório e Fundamentação de Facto podem, de facto, complementar a secção de Fundamentação de Direito na definição do espaço de descritores.

Em algumas secções dentro do domínio Cível, a configuração **todas as secções/todos os modelos**, que utiliza o documento completo como espaço de embeddings, obteve o melhor desempenho. No entanto, a diferença de desempenho em relação às configurações **l, r-l** e **r-f-l** é mínima. Uma vantagem de utilizar apenas a secção de Fundamentação de Direito como espaço de embeddings, que apresentou o melhor desempenho em várias secções, é a redução significativa do tamanho dos acórdãos, que tendem a ser extensos. Do ponto de vista computacional, esta abordagem pode diminuir o tempo de processamento e os recursos computacionais necessários.

Todos os melhores resultados, com exceção de um na Secção 1, foram alcançados utilizando **todos os modelos da área Cível** para atribuir os descritores. Conforme descrito na Figura 5, cada modelo da respetiva área prevê os descritores para cada acórdão do conjunto de teste dentro de uma secção dessa área. As pontuações de cada descritor são somadas, e os descritores com as pontuações mais altas são retornados. Esta abordagem beneficia os descritores que estão presentes em várias secções. No entanto, as pontuações só são atribuídas a descritores que pertencem ao espaço de descritores do acórdão em questão. Dessa forma, esta abordagem favorece os descritores que aparecem no espaço de descritores do acórdão e que estão distribuídos por várias secções.

As Tabelas 8 e 9 apresentam os resultados para as secções dentro da Área Criminal, assim como para a Secção Quatro (Área Social) e para a Área Contencioso, respetivamente. De forma consistente com as secções do domínio Cível, os melhores desempenhos são observados nas configurações onde a secção de Fundamentação de Direito está incluída. No domínio Criminal, apenas as configurações **r-f-l** e **todas as secções** e **todas as secções/todos os modelos** alcançaram os melhores desempenhos. No entanto, configurações como **l** e **r-l** obtiveram resultados muito semelhantes. A configuração **d** continua a ser a pior em termos de desempenho; contudo, exceto na Secção Quatro, os resultados estão mais próximos dos melhores desempenhos do que aqueles obtidos no domínio Cível. Para o domínio Criminal, quase todos os melhores desempenhos são alcançados quando se utiliza o acórdão completo. No entanto, tal como no domínio Cível, dado que a diferença de desempenho é mínima,

| | Secções | p/r@1 | p/r@2 | p/r@3 | p/r@4 | p/r@5 | p/r@10 | p/r@15 | p/r@20 |
|---|---|---|---|---|---|---|---|---|---|
| Secção 1 | r | 0,51 | 0,44 | 0,41 | 0,4 | 0,4 | 0,45 | 0,51 | 0,55 |
| | r todos os modelos | 0,43 | 0,37 | 0,35 | 0,34 | 0,35 | 0,4 | 0,45 | 0,5 |
| | l | 0,51 | 0,43 | 0,4 | 0,4 | 0,4 | 0,45 | 0,5 | 0,54 |
| | l todos os modelos | 0,51 | 0,44 | 0,42 | 0,41 | 0,41 | 0,47 | 0,53 | 0,57 |
| | f-l | 0,5 | 0,43 | 0,39 | 0,39 | 0,39 | 0,44 | 0,49 | 0,52 |
| | f-l todos os modelos | 0,49 | 0,43 | 0,41 | 0,4 | 0,4 | 0,46 | 0,51 | 0,57 |
| | r-f-l | 0,51 | 0,44 | 0,41 | 0,4 | 0,4 | 0,45 | 0,51 | 0,55 |
| | r-f-l todos os modelos | 0,5 | 0,45 | **0,43** | **0,42** | 0,42 | **0,48** | **0,54** | **0,58** |
| | r-l | 0,51 | 0,44 | 0,41 | 0,41 | 0,41 | 0,46 | 0,51 | 0,55 |
| | r-l todos os modelos | 0,51 | 0,45 | **0,43** | **0,42** | 0,42 | 0,47 | 0,53 | 0,57 |
| | d | 0,12 | 0,1 | 0,09 | 0,09 | 0,09 | 0,1 | 0,12 | 0,14 |
| | d todos os modelos | 0,12 | 0,11 | 0,1 | 0,1 | 0,1 | 0,12 | 0,15 | 0,17 |
| | r-l-d | 0,37 | 0,32 | 0,3 | 0,29 | 0,29 | 0,33 | 0,37 | 0,41 |
| | r-l-d todos os modelos | 0,39 | 0,34 | 0,32 | 0,31 | 0,32 | 0,36 | 0,42 | 0,47 |
| | todas as secções | **0,52** | 0,45 | 0,41 | 0,41 | 0,41 | 0,46 | 0,52 | 0,55 |
| | todas as secções/todos os modelos | **0,52** | **0,46** | **0,43** | **0,42** | **0,43** | **0,48** | **0,54** | **0,58** |
| Secção 2 | r | 0,42 | 0,35 | 0,33 | 0,32 | 0,31 | 0,34 | 0,39 | 0,43 |
| | r todos os modelos | 0,43 | 0,38 | 0,35 | 0,34 | 0,34 | 0,38 | 0,44 | 0,49 |
| | l | 0,49 | 0,44 | 0,4 | 0,39 | 0,38 | 0,43 | 0,47 | 0,5 |
| | l todos os modelos | **0,54** | **0,47** | 0,43 | **0,42** | **0,42** | 0,45 | 0,51 | 0,55 |
| | f-l | 0,49 | 0,42 | 0,39 | 0,37 | 0,37 | 0,4 | 0,46 | 0,5 |
| | f-l todos os modelos | 0,52 | 0,45 | 0,42 | 0,4 | 0,4 | 0,44 | 0,49 | 0,54 |
| | r-f-l | 0,49 | 0,43 | 0,4 | 0,39 | 0,38 | 0,43 | 0,48 | 0,52 |
| | r-f-l todos os modelos | 0,52 | 0,45 | 0,43 | 0,41 | **0,42** | **0,46** | **0,52** | **0,57** |
| | r-l | 0,51 | 0,43 | 0,4 | 0,39 | 0,39 | 0,43 | 0,48 | 0,52 |
| | r-l todos os modelos | **0,54** | 0,46 | 0,43 | **0,42** | 0,41 | **0,46** | **0,52** | 0,56 |
| | d | 0,12 | 0,12 | 0,11 | 0,1 | 0,11 | 0,11 | 0,13 | 0,15 |
| | d todos os modelos | 0,14 | 0,13 | 0,13 | 0,12 | 0,12 | o.13 | 0,16 | 0,18 |
| | r-l-d | 0,38 | 0,34 | 0,32 | 0,3 | 0,3 | 0,33 | 0,37 | 0,41 |
| | r-l-d todos os modelos | 0,39 | 0,36 | 0,32 | 0,32 | 0,32 | 0,36 | 0,42 | 0,47 |
| | todas as secções | 0,49 | 0,44 | 0,4 | 0,39 | 0,39 | 0,43 | 0,48 | 0,52 |
| | todas as secções/todos os modelos | 0,53 | 0,46 | **0,44** | **0,42** | **0,42** | **0,46** | **0,52** | **0,57** |
| Secção 6 | r | 0,42 | 0,37 | 0,34 | 0,32 | 0,32 | 0,36 | 0,41 | 0,44 |
| | r todos os modelos | 0,44 | 0,4 | 0,35 | 0,34 | 0,35 | 0,4 | 0,47 | 0,51 |
| | l | 0,52 | 0,44 | 0,41 | 0,39 | 0,39 | 0,44 | 0,49 | 0,53 |
| | l todos os modelos | 0,52 | 0,47 | 0,43 | 0,41 | 0,41 | **0,49** | **0,55** | 0,59 |
| | f-l | 0,5 | 0,43 | 0,39 | 0,37 | 0,36 | 0,42 | 0,48 | 0,52 |
| | f-l todos os modelos | 0,5 | 0,46 | 0,42 | 0,4 | 0,4 | 0,47 | 0,54 | 0,58 |
| | r-f-l | 0,5 | 0,44 | 0,4 | 0,39 | 0,38 | 0,44 | 0,49 | 0,52 |
| | r-f-l todos os modelos | 0,51 | 0,47 | **0,44** | 0,42 | **0,42** | 0,48 | **0,55** | 0,6 |
| | r-l | 0,51 | 0,44 | 0,41 | 0,39 | 0,38 | 0,43 | 0,49 | 0,55 |
| | r-l todos os modelos | 0,52 | **0,48** | **0,44** | 0,42 | **0,42** | **0,49** | **0,55** | **0,61** |
| | d | 0,13 | 0,11 | 0,1 | 0,09 | 0,09 | 0,11 | 0,14 | 0,15 |
| | d todos os modelos | 0,15 | 0,14 | 0,13 | 0,12 | 0,12 | 0,15 | 0,18 | 0,2 |
| | r-l-d | 0,46 | 0,4 | 0,38 | 0,36 | 0,36 | 0,41 | 0,46 | 0,49 |
| | r-l-d todos os modelos | 0,49 | 0,45 | 0,4 | 0,39 | 0,38 | 0,45 | 0,51 | 0,56 |
| | todas as secções | 0,51 | 0,44 | 0,41 | 0,39 | 0,39 | 0,44 | 0,5 | 0,53 |
| | todas as secções/todos os modelos | **0,53** | **0,48** | **0,44** | **0,43** | **0,42** | **0,49** | **0,55** | 0,6 |
| Secção 7 | r | 0,41 | 0,36 | 0,34 | 0,33 | 0,32 | 0,34 | 0,39 | 0,42 |
| | r todos os modelos | 0,45 | 0,39 | 0,37 | 0,36 | 0,35 | 0,38 | 0,44 | 0,49 |
| | l | 0,53 | 0,46 | 0,43 | 0,4 | 0,4 | 0,42 | 0,46 | 0,5 |
| | l todos os modelos | **0,58** | **0,51** | **0,46** | **0,43** | **0,43** | 0,45 | **0,51** | 0,55 |
| | f-l | 0,51 | 0,44 | 0,4 | 0,39 | 0,38 | 0,4 | 0,45 | 0,49 |
| | f-l todos os modelos | 0,55 | 0,48 | 0,45 | 0,42 | 0,42 | 0,44 | 0,5 | 0,54 |
| | r-f-l | 0,51 | 0,46 | 0,42 | 0,4 | 0,39 | 0,42 | 0,46 | 0,5 |
| | r-f-l todos os modelos | 0,53 | 0,5 | **0,46** | **0,43** | 0,42 | **0,46** | **0,51** | 0,55 |
| | r-l | 0,51 | 0,45 | 0,41 | 0,39 | 0,39 | 0,41 | 0,46 | 0,5 |
| | r-l todos os modelos | 0,55 | 0,5 | 0,45 | **0,43** | 0,42 | 0,45 | **0,51** | **0,56** |
| | d | 0,14 | 0,13 | 0,11 | 0,11 | 0,1 | 0,11 | 0,13 | 0,14 |
| | d todos os modelos | 0,14 | 0,13 | 0,12 | 0,11 | 0,11 | 0,13 | 0,15 | 0,18 |
| | r-l-d | 0,46 | 0,41 | 0,38 | 0,35 | 0,35 | 0,37 | 0,42 | 0,45 |
| | r-l-d todos os modelos | 0,51 | 0,43 | 0,4 | 0,38 | 0,38 | 0,42 | 0,47 | 0,51 |
| | todas as secções | 0,52 | 0,46 | 0,43 | 0,4 | 0,4 | 0,42 | 0,46 | 0,5 |
| | todas as secções/todos os modelos | 0,55 | 0,5 | 0,45 | **0,43** | **0,43** | **0,46** | **0,51** | **0,56** |
| Área Cível | r | 0,45 | 0,39 | 0,36 | 0,34 | 0,33 | 0,36 | 0,41 | 0,44 |
| | l | **0,56** | **0,48** | **0,44** | **0,42** | 0,4 | **0,44** | **0,49** | 0,52 |
| | f-l | 0,53 | 0,46 | 0,42 | 0,4 | 0,4 | 0,42 | 0,47 | 0,51 |
| | r-f-l | 0,53 | 0,47 | 0,43 | 0,41 | 0,4 | 0,43 | 0,48 | 0,52 |
| | r-l | 0,55 | **0,48** | **0,44** | 0,41 | **0,41** | **0,44** | **0,49** | 0,52 |
| | d | 0,14 | 0,12 | 0,11 | 0,1 | 0,1 | 0,11 | 0,13 | 0,14 |
| | r-l-d | 0,5 | 0,43 | 0,4 | 0,38 | 0,37 | 0,4 | 0,45 | 0,48 |
| | todas as secções | 0,54 | 0,47 | 0,43 | 0,41 | 0,4 | **0,44** | **0,49** | **0,53** |

**Tabela 7:** Resultados utilizando a abordagem SLEEC e considerando diferentes secções dos acórdãos para as secções dentro da Área Cível.

| | Secções | p/r@1 | p/r@2 | p/r@3 | p/r@4 | p/r@5 | p/r@10 | p/r@15 | p/r@20 |
|---|---|---|---|---|---|---|---|---|---|
| Secção 3 | r | 0,51 | 0,45 | 0,41 | 0,39 | 0,38 | 0,41 | 0,46 | 0,5 |
| | r todos os modelos | 0,51 | 0,45 | 0,41 | 0,39 | 0,38 | 0,42 | 0,47 | 0,51 |
| | l | 0,58 | 0,51 | 0,47 | 0,46 | 0,45 | 0,48 | 0,53 | 0,57 |
| | l todos os modelos | 0,57 | 0,52 | 0,47 | 0,45 | 0,44 | 0,48 | 0,54 | 0,58 |
| | f-l | 0,59 | 0,51 | 0,47 | 0,46 | 0,45 | 0,48 | 0,53 | 0,56 |
| | f-l todos os modelos | 0,59 | 0,52 | 0,47 | 0,45 | 0,44 | 0,48 | 0,54 | 0,58 |
| | r-f-l | **0,6** | **0,53** | **0,49** | 0,47 | 0,45 | 0,49 | 0,54 | 0,58 |
| | r-f-l todos os modelos | 0,58 | 0,52 | 0,48 | 0,46 | 0,45 | 0,49 | 0,54 | 0,59 |
| | r-l | 0,59 | 0,53 | 0,48 | 0,46 | 0,46 | 0,49 | 0,54 | 0,58 |
| | r-l todos os modelos | 0,58 | 0,51 | 0,47 | 0,45 | 0,45 | 0,49 | 0,53 | 0,59 |
| | d | 0,43 | 0,36 | 0,33 | 0,31 | 0,31 | 0,33 | 0,36 | 0,39 |
| | d todos os modelos | 0,44 | 0,38 | 0,35 | 0,33 | 0,32 | 0,35 | 0,38 | 0,42 |
| | r-l-d | 0,54 | 0,46 | 0,44 | 0,42 | 0,41 | 0,43 | 0,47 | 0,51 |
| | r-l-d todos os modelos | 0,55 | 0,47 | 0,43 | 0,42 | 0,41 | 0,44 | 0,49 | 0,53 |
| | todas as secções | **0,6** | **0,53** | **0,49** | **0,48** | **0,47** | **0,5** | 0,55 | 0,59 |
| | todas as secções/todos os modelos | **0,6** | **0,53** | **0,49** | 0,47 | 0,46 | **0,5** | **0,56** | **0,61** |
| Secção 5 | r | 0,52 | 0,45 | 0,41 | 0,4 | 0,39 | 0,43 | 0,48 | 0,53 |
| | r todos os modelos | 0,55 | 0,47 | 0,42 | 0,41 | 0,4 | 0,45 | 0,51 | 0,56 |
| | l | 0,58 | 0,51 | 0,47 | 0,46 | 0,45 | 0,49 | 0,55 | 0,59 |
| | l todos os modelos | 0,6 | 0,52 | 0,48 | 0,46 | 0,45 | 0,49 | 0,57 | 0,62 |
| | f-l | 0,62 | 0,53 | 0,48 | 0,46 | 0,45 | 0,49 | 0,54 | 0,58 |
| | f-l todos os modelos | 0,61 | 0,53 | 0,48 | 0,46 | 0,45 | 0,5 | 0,57 | 0,62 |
| | r-f-l | 0,62 | 0,53 | 0,49 | 0,47 | 0,46 | 0,49 | 0,55 | 0,59 |
| | r-f-l todos os modelos | 0,61 | 0,52 | 0,48 | 0,47 | **0,47** | 0,51 | 0,58 | 0,62 |
| | r-l | 0,61 | 0,53 | 0,49 | 0,46 | 0,45 | 0,5 | 0,55 | 0,59 |
| | r-l todos os modelos | 0,62 | 0,52 | 0,49 | 0,47 | **0,47** | 0,51 | 0,58 | 0,62 |
| | d | 0,44 | 0,37 | 0,34 | 0,32 | 0,32 | 0,36 | 0,4 | 0,43 |
| | d todos os modelos | 0,44 | 0,37 | 0,34 | 0,32 | 0,32 | 0,37 | 0,42 | 0,46 |
| | r-l-d | 0,56 | 0,46 | 0,43 | 0,41 | 0,4 | 0,44 | 0,49 | 0,53 |
| | r-l-d todos os modelos | 0,57 | 0,49 | 0,45 | 0,43 | 0,41 | 0,47 | 0,53 | 0,58 |
| | todas as secções | 0,63 | **0,54** | **0,5** | **0,48** | **0,47** | 0,5 | 0,56 | 0,6 |
| | todas as secções/todos os modelos | **0,64** | **0,54** | 0,49 | **0,48** | **0,47** | **0,52** | **0,59** | **0,63** |
| Área Criminal | r | 0,55 | 0,48 | 0,43 | 0,41 | 0,4 | 0,43 | 0,48 | 0,52 |
| | l | **0,6** | 0,52 | 0,47 | 0,44 | 0,44 | 0,48 | 0,53 | 0,57 |
| | f-l | 0,59 | 0,51 | 0,47 | 0,45 | 0,44 | 0,48 | 0,53 | 0,57 |
| | r-f-l | **0,6** | **0,53** | 0,48 | **0,46** | 0,45 | 0,49 | **0,55** | **0,59** |
| | r-l | **0,6** | 0,52 | 0,48 | 0,45 | 0,44 | 0,48 | 0,53 | 0,57 |
| | d | 0,43 | 0,36 | 0,33 | 0,31 | 0,3 | 0,33 | 0,36 | 0,4 |
| | r-l-d | 0,57 | 0,5 | 0,45 | 0,43 | 0,42 | 0,46 | 0,51 | 0,55 |
| | todas as secções | **0,6** | **0,53** | **0,49** | **0,46** | **0,46** | **0,5** | **0,55** | **0,59** |

**Tabela 8:** Resultados utilizando a abordagem SLEEC e considerando diferentes secções dos acórdãos para as secções dentro da Área Criminal.

é mais vantajoso utilizar configurações que reduzam o tamanho do acórdão devido aos custos computacionais. Ao contrário do domínio Cível, no domínio Criminal, os melhores desempenhos são obtidos ao utilizar apenas o modelo da respetiva secção. Dado o elevado grau de sobreposição de descritores entre as secções Três e Cinco, como mostrado na Tabela 5, seria natural assumir que a combinação dos modelos melhoraria o desempenho, pelos mesmos motivos explicados para o domínio Cível.

O domínio Cível obteve os piores resultados em comparação com os domínios Criminal, Social e Contencioso. Conforme mostrado na Tabela 4, as secções dentro do domínio Cível possuem um número significativamente maior de descritores, o que justifica os resultados mais baixos, uma vez que o número de descritores possíveis a serem atribuídos é maior. A Área Contencioso apresenta um número muito reduzido de descritores, o que explica o seu desempenho superior em comparação com as outras áreas. No entanto, os resultados não são tão elevados quanto poderiam ser, pois, ao contrário das outras áreas, a Área Contencioso tem um número de descritores superior ao número de acórdãos disponíveis,

| | Secções | p/r@1 | p/r@2 | p/r@3 | p/r@4 | p/r@5 | p/r@10 | p/r@15 | p/r@20 |
|---|---|---|---|---|---|---|---|---|---|
| Secção 4 | r | 0,47 | 0,42 | 0,4 | 0,4 | 0,43 | 0,5 | 0,54 | 0,56 |
| | l | 0,55 | 0,5 | **0,48** | **0,48** | **0,5** | **0,58** | **0,63** | **0,66** |
| | f-l | 0,54 | 0,48 | 0,44 | 0,45 | 0,46 | 0,56 | 0,61 | 0,64 |
| | r-f-l | 0,55 | 0,48 | 0,46 | 0,46 | 0,49 | 0,56 | 0,61 | 0,64 |
| | r-l | 0,55 | **0,51** | **0,48** | **0,48** | **0,5** | **0,58** | **0,63** | 0,65 |
| | d | 0,19 | 0,16 | 0,16 | 0,16 | 0,17 | 0,2 | 0,23 | 0,26 |
| | r-l-d | 0,49 | 0,41 | 0,38 | 0,39 | 0,41 | 0,49 | 0,53 | 0,56 |
| | all sections | **0,56** | 0,49 | 0,46 | 0,47 | 0,49 | 0,57 | 0,62 | 0,65 |
| Contencioso | r | 0,62 | 0,52 | 0,49 | 0,46 | 0,44 | 0,41 | 0,46 | 0,5 |
| | l | **0,68** | 0,61 | 0,57 | 0,54 | 0,51 | 0,48 | **0,53** | 0,56 |
| | f-l | **0,68** | **0,64** | 0,56 | 0,55 | **0,53** | **0,5** | 0,51 | 0,56 |
| | r-f-l | 0,64 | 0,63 | **0,58** | **0,56** | 0,52 | 0,46 | 0,52 | **0,59** |
| | r-l | 0,59 | 0,53 | 0,53 | 0,51 | 0,49 | 0,46 | 0,5 | 0,52 |
| | d | 0,56 | 0,44 | 0,41 | 0,37 | 0,36 | 0,31 | 0,34 | 0,38 |
| | r-l-d | 0,55 | 0,48 | 0,44 | 0,41 | 0,4 | 0,39 | 0,44 | 0,48 |
| | all sections | 0,59 | 0,6 | 0,56 | 0,52 | **0,53** | 0,48 | 0,52 | 0,57 |

**Tabela 9:** Resultados utilizando a abordagem SLEEC e considerando diferentes secções dos acórdãos para as secções dentro das Áreas Social e Contencioso.

como demonstrado na Tabela 4. As secções das áreas Criminal e Social também beneficiam de um menor número de descritores em comparação com a área Cível. Por exemplo, a Secção 1 tem mais do que o dobro do número de descritores da Secção 5. Na Secção 3, levantámos a hipótese de que separar as áreas em secções para reduzir o número de descritores melhoraria os resultados. Como evidenciado nas Tabelas 7 e 8, utilizar secções geralmente conduz a melhores resultados do que utilizar áreas, em todas as variações de precisão/cobertura. Conforme esperado, as métricas **p/r@15** e **p/r@20** alcançaram os melhores resultados para quase todas as secções e áreas. Dado que estas duas variações de precisão/cobertura retornam um maior número de descritores e que, tipicamente, o número de descritores associados a um acórdão é inferior a 15, há uma maior probabilidade de os descritores retornados corresponderem aos descritores corretos. No entanto, apesar de os resultados poderem ser significativamente melhorados com valores mais altos de precisão/cobertura, os resultados obtidos com **p/r@1** são bastante satisfatórios. Num intervalo de 439 a 1524 descritores, todos os modelos alcançaram um resultado superior a 0,5 ao retornarem apenas um descritor que está corretamente associado ao acórdão testado.

## 6. Conclusão e Trabalho Futuro

Neste artigo, explorámos a tarefa de atribuição automática de descritores aos acórdãos do Supremo Tribunal de Justiça de Portugal. Esta investigação foi realizada no âmbito do projeto *IRIS*, uma colaboração entre o Supremo Tribunal de Justiça de Portugal e o INESC-ID. A abordagem utilizada baseia-se em representações locais esparsas, SLEEC (Bhatia et al., 2015), que apresentaram previamente bons resultados em conjuntos de dados jurídicos em língua Inglesa. Tanto quanto nos é possível saber, este é o primeiro esforço para atribuir automaticamente descritores oficiais do Supremo Tribunal de Justiça de Portugal aos seus acórdãos.

A nossa investigação incluiu testar a abordagem SLEEC em dados jurídicos do Português Europeu. Além disso, a estruturação dos acórdãos em secções para mapear diferentes espaços de *embeddings* permitiu-nos reduzir o tamanho dos acórdãos, diminuindo assim os custos computacionais, e compreender a importância das diferentes secções na definição dos descritores-chave de um acórdão.

Os nossos resultados experimentais refletem a complexidade inerente à tarefa, que decorre da dimensão limitada do conjunto de dados e da elevada dimensionalidade do espaço de saída (grande número de etiquetas), ambas características do problema XML. Apesar destes desafios, a abordagem proposta apresentou resulta-

dos promissores na classificação de documentos de acórdãos jurídicos em Português Europeu.

Para trabalho futuro, propomos várias direções para melhorar o desempenho da atribuição de descritores neste tipo de contextos. Primeiramente, aumentar o tamanho do conjunto de treino poderá permitir que o modelo aprenda melhor os padrões dos acórdãos, melhorando a precisão da atribuição de descritores. Além disso, a investigação de técnicas de representações de texto baseadas em *embeddings* mais avançadas, incluindo modelos baseados em aprendizagem profunda, poderá capturar representações mais sofisticadas dos textos jurídicos.

Em particular, planeamos explorar abordagens recentes baseadas em *transformers* aplicadas à XML, como o *LightXML* (Jiang et al., 2021), o *DeepXML* (Zhang et al., 2017) e o *AttentionXML* (You et al., 2018), que têm demonstrado resultados promissores em conjuntos de dados de grande escala, o que permitirá obter uma perspetiva mais clara sobre as suas vantagens e limitações em cenários com dados no domínio legal em língua portuguesa e com um espaço de saída altamente não balanceado.

Por fim, uma análise mais aprofundada da contribuição de cada secção do acórdão na atribuição de descritores poderá levar ao desenvolvimento de modelos mais eficientes, possivelmente através da atribuição de pesos diferenciados às secções com maior relevância.

## Agradecimentos

Esta investigação foi apoiada pela Fundação para a Ciência e Tecnologia (FCT), através do financiamento plurianual do INESC-ID, com a referência DOI:10.54499/UIDB/50021/2020. No INESC-ID, esta investigação está inserida no âmbito do projeto IRIS, uma colaboração com o Supremo Tribunal de Justiça de Portugal, com a referência interna PR07005. Esta investigação foi igualmente apoiada pelo Plano de Recuperação e Resiliência de Portugal, através do projeto C645008882-00000055.

## Referências

Agrawal, Rahul, Archit Gupta, Yashoteja Prabhu & Manik Varma. 2013. Multi-label learning with millions of labels: recommending advertiser bid phrases for web pages. Em *22nd International World Wide Web Conference (WWW)*, 13–24. 10.1145/2488388.2488391

Babbar, Rohit & Bernhard Schölkopf. 2017. DiSMEC: Distributed sparse machines for extreme multi-label classification. Em *10th International Conference on Web Search and Data Mining (WSDM)*, 721–729. 10.1145/3018661.3018741

Babbar, Rohit & Bernhard Schölkopf. 2018. Adversarial extreme multi-label classification. arXiv [stat.ML/cs.LG]. 10.48550/arXiv.1803.01570

Bhatia, Kush, Himanshu Jain, Purushottam Kar, Manik Varma & Prateek Jain. 2015. Sparse local embeddings for extreme multi-label classification. Em *Advances in Neural Information Processing Systems (NIPS)*, 730–738.

Bi, Wei & James Tin-Yau Kwok. 2013. Efficient multi-label classification with many labels. Em *30th International Conference on Machine Learning (ICML)*, 405–413.

Boutsidis, Christos, Michael W. Mahoney & Petros Drineas. 2009. An improved approximation algorithm for the column subset selection problem. Em *20th Symposium on Discrete Algorithms (SoDA)*, 968–977. 10.1137/1.9781611973068.105

Cissé, Moustapha, Nicolas Usunier, Thierry Artières & Patrick Gallinari. 2013. Robust bloom filters for large multilabel classification tasks. Em *Advances in Neural Information Processing Systems (NIPCS)*, 1851–1859.

Dong, Wei, Moses Charikar & Kai Li. 2011. Efficient k-nearest neighbor graph construction for generic similarity measures. Em *20th International Conference on World Wide Web (WWW)*, 577–586. 10.1145/1963405.1963487

Fang, Huang, Minhao Cheng, Cho-Jui Hsieh & Michael P. Friedlander. 2019. Fast training for large-scale one-versus-all linear classifiers using tree-structured initialization. Em *International Conference on Data Mining (SDM)*, 280–288. 10.1137/1.9781611975673.32

Jain, Himanshu, Yashoteja Prabhu & Manik Varma. 2016. Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications. Em *22nd International Conference on Knowledge Discovery and Data Mining (KDD)*, 935–944. 10.1145/2939672.2939756

Jain, Prateek, Raghu Meka & Inderjit S. Dhillon. 2010. Guaranteed rank minimization via singular value projection. Em *Advances in Neural Information Processing Systems (NIPS)*, 937–945.

Jiang, Ting, Deqing Wang, Leilei Sun, Huayi Yang, Zhengyang Zhao & Fuzhen Zhuang. 2021. LightXML: Transformer with dynamic negative sampling for high-performance extreme multi-label text classification. Em *35th Conference on Artificial Intelligence (AAAI)*, 7987–7994. 10.1609/aaai.V35I9.16974

Kanapala, Ambedkar, Sukomal Pal & Rajendra Pamula. 2019. Text summarization from legal documents: a survey. *Artificial Intelligence Review* 51(3). 371–402. 10.1007/s10462-017-9566-2

Khandagale, Sujay, Han Xiao & Rohit Babbar. 2019. Bonsai - diverse and shallow trees for extreme multi-label classification. arXiv [cs.LG/stat.ML]. 10.48550/arXiv.1904.08249

Medini, Tharun, Qixuan Huang, Yiqiu Wang, Vijai Mohan & Anshumali Shrivastava. 2019. Extreme classification in log memory using count-min sketch: A case study of amazon search with 50m products. Em *Advances in Neural Information Processing Systems (NIPS)*, 13244–13254.

Melo, Rui, Pedro A. Santos & João Dias. 2023. A semantic search system for the supremo tribunal de justiça. Em *22nd Portuguese Conference on Artificial Intelligence (EPIA)*, 142–154. 10.1007/978-3-031-49011-8_12

Niculescu-Mizil, Alexandru & Ehsan Abbasnejad. 2017. Label filters for large scale multilabel classification. Em *20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 1448–1457.

Papanikolaou, Yannis, Ioannis Katakis & Grigorios Tsoumakas. 2016. Hierarchical partitioning of the output space in multi-label data. arXiv [stat.ML/cs.LG]. 10.48550/arXiv.1612.06083

Pennington, Jeffrey, Richard Socher & Christopher Manning. 2014. GloVe: Global vectors for word representation. Em *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. 10.3115/v1/D14-1162

Prabhu, Yashoteja & Manik Varma. 2014. FastXML: a fast, accurate and stable tree-classifier for extreme multi-label learning. Em *International Conference on Knowledge Discovery and Data Mining, (KDD)*, 263–272. 10.1145/2623330.2623651

Souza, Fábio, Rodrigo Frassetto Nogueira & Roberto de Alencar Lotufo. 2020. BERTimbau: Pretrained BERT models for Brazilian Portuguese. Em *9th Brazilian Conference on Intelligent Systems (BRACIS)*, 403–417. 10.1007/978-3-030-61377-8_28

Tagami, Yukihiro. 2017. AnnexML: Approximate nearest neighbor search for extreme multi-label classification. Em *International Conference on Knowledge Discovery and Data Mining (KDD)*, 455–464. 10.1145/3097983.3097987

Tai, Farbound & Hsuan-Tien Lin. 2012. Multilabel classification with principal label space transformation. *Neural Computation* 24(9). 2508–2542. 10.1162/NECO_A_00320

Tsoumakas, Grigorios, Ioannis Katakis & Ioannis Vlahavas. 2008. Effective and efficient multilabel classification in domains with large number of labels. Em *Workshop on Mining Multidimensional Data (MMD)*, 53–59.

Turtle, Howard. 1995. Text retrieval in the legal world. *Artificial Intelligence and Law* 3(1). 5–54. 10.1007/BF00877694

Wang, Bingyu, Li Chen, Wei Sun, Kechen Qin, Kefeng Li & Hui Zhou. 2019. Ranking-based autoencoder for extreme multi-label classification. arXiv [cs.LG/stat.ML]. 10.48550/arXiv.1904.05937

Wei, Tong, Zhen Mao, Jiang-Xin Shi, Yu-Feng Li & Min-Ling Zhang. 2022. A survey on extreme multi-label learning. arXiv [cs.LG]. 10.48550/arXiv.2210.03968

Yen, Ian En-Hsu, Xiangru Huang, Pradeep Ravikumar, Kai Zhong & Inderjit S. Dhillon. 2016. PD-Sparse : A primal and dual sparse approach to extreme multiclass and multilabel classification. Em *33rd International Conference on Machine Learning (ICML)*, 3069–3077.

You, Ronghui, Suyang Dai, Zihan Zhang, Hiroshi Mamitsuka & Shanfeng Zhu. 2018. AttentionXML: Extreme multi-label text classification with multi-label attention based recurrent neural networks. arXiv [cs.CL/cs.LG]. 10.48550/arXiv.1811.01727

Zanatti, Martim, Ricardo Ribeiro & H. Sofia Pinto. 2024. Segmenting model for portuguese judgments. Em *Portuguese Conference on Artificial Intelligence (EPIA)*, 245–257. 10.1007/978-3-031-73497-7_20

Zhang, Wenjie, Liwei Wang, Junchi Yan, Xiangfeng Wang & Hongyuan Zha. 2017. Deep extreme multi-label learning. arXiv [cs.LG]. 10.48550/arXiv.1704.03718