

Reconhecimento de entidades nomeadas em documentos jurídicos em português utilizando redes neurais.

Caio C. R. Mota¹, André C.A. Nascimento¹, Péricles B.C. Miranda¹,
Rafael Ferreira Mello¹, Isabel W. S. Maldonado², José L. M. Coelho Filho²

¹ Departamento de Computação, Universidade Federal
Rural de Pernambuco (UFRPE), Recife, Brasil
² NESS Law, São Paulo, Brasil

Abstract. *Over the past few years, information technology has been transforming the legal world, automating processes and, consequently, reducing the time needed to create and analyse digital legal documents. One of the most studied problems in this area is the recognition of named entities (NER) in unstructured texts. Previous work has not addressed the detection of legal entities through the application of models based on neural networks available in natural language processing libraries. In this article, the use of the libraries Spacy and FLAIR were analyzed in the context of NER in initial petitions. The models were trained with pre-defined architectures and evaluated in two corpora, one of them developed in the scope of this work. The results obtained with these experiments demonstrated good results with both platforms Spacy and FLAIR, with superior performance when adopting BiLSTM-CRF with FLAIR embeddings.*

Resumo. *Ao longo dos últimos anos, a tecnologia da informação vem transformando o mundo jurídico, automatizando processos e, por consequência, diminuindo o tempo necessário para criação e análise de peças jurídicas digitais. Um dos problemas mais estudados nesta área é o reconhecimento de entidades nomeadas (REN) em textos não estruturados. Trabalhos anteriores não abordaram a detecção de entidades legais por meio da aplicação de modelos baseados em redes neurais disponíveis em bibliotecas de processamento de linguagens natural. Neste artigo, o uso de das bibliotecas Spacy e FLAIR foram analisados no contexto de REN em petições iniciais. Os modelos foram treinados com arquiteturas pré-definidas e avaliados em dois corpora, um deles desenvolvido no âmbito deste trabalho. Os resultados obtidos com esses experimentos demonstraram bons resultados com ambas as plataformas Spacy e FLAIR, com desempenho superior quando adotado o BiLSTM-CRF com FLAIR embeddings.*

1. Introdução

A tecnologia da informação vem, ao longo dos últimos anos, transformando o mundo jurídico, motivando a criação de novos negócios e soluções tecnológicas com grande velocidade. A Associação Brasileira de *Lawtechs* e *Legaltechs* (AB2L)¹, por exemplo, registra que de 2017 a 2019 o setor jurídico já conta com quase 150 *startups* [D'Angelo 2019]. Além do ganho de tempo na criação de peças jurídicas propiciado pelos meios digitais, a popularização de aplicações de inteligência artificial está abrindo caminho para o ganho de informações, auxiliando na tomada de decisões dos agentes jurídicos.

¹<https://ab2l.org.br/>

A inteligência artificial vem auxiliando o meio jurídico através de técnicas já estabelecidas, principalmente na área de processamento de linguagens natural (PLN) [Dale 2019]. A maioria dos dados jurídicos produzidos dentro e fora dos tribunais são definidos em textos livres em sistemas, ou ainda, em documentos não-estruturados (e.g., em formato PDF), ou seja, as informações apresentadas não estão organizadas ou ordenadas de forma padronizada. Dessa forma, tais informações não são acessíveis a modelos tradicionais de processamento, recuperação e extração de informações. Nesse contexto, existem vários algoritmos de PLN que tem sido aplicados neste tipo de sistema, com o objetivo de estruturar esses dados, a fim de ajudar no processo de tomada de decisão, bem como agilizar o processo judicial [Dale 2019].

Neste sentido, este trabalho propõe uma avaliação de técnicas de extração de informação chamada de reconhecimento de entidades nomeadas (do inglês, *Named Entity Recognition* - NER) [Yadav and Bethard 2019] para o contexto jurídico brasileiro. Estas técnicas consistem em localizar em um texto palavras e classificá-las de acordo com rótulos previamente definidos. Aplicações comuns de NER consistem em extrair nomes de pessoas, locais, valores monetários e até datas [Minkov et al. 2005]. No contexto jurídico, as entidades já estudadas incluem leis [Luz de Araujo et al. 2018, Chen et al. 2020], reconhecimento de partes lógicas em documentos [Son et al. 2016], pessoa, jurisprudência, legislação, tempo, localização e organização [Luz de Araujo et al. 2018].

Trabalhos como o de [Luz de Araujo et al. 2018], [Chen et al. 2020] e [Yadav and Bethard 2019] focaram na experimentação de técnicas já estabelecidas no estado-da-arte em cima das bases de dados propostas, as quais eram o artefato central das pesquisas. Esta pesquisa, por outro lado, objetiva-se não só a experimentação de modelos do estado-da-arte na base de dados proposta, mas também a análise do desempenho de bibliotecas de PLN e seus respectivos *frameworks*. Para mais, é explorado o efeito do uso de *Flair embeddings* em conjunto com um modelo BiLSTM-CRF no contexto específico de textos jurídicos, onde pode-se analisar o desempenho dessa arquitetura em um *dataset* de domínio específico como o sugerido por este trabalho. Estes resultados são importantes para que se tenha uma ideia mais profunda da amplitude de aplicação de métodos de PLN na resolução de problema do setor jurídico. O presente trabalho se propõe a extrair legislações e jurisprudências como entidades nomeadas de petições iniciais. Duas bases de dados foram avaliadas: a base de dados LeNER-br, proposta originalmente por [Luz de Araujo et al. 2018] e uma base de dados proprietária, produzida neste trabalho, a qual consiste de petições iniciais de processos que tramitaram na primeira instância da justiça Brasileira. Esta base é composta por 1676 trechos de petições iniciais, contendo um total de 3825 anotações das entidades legislação e jurisprudência. Vale-se ressaltar ainda que [Luz de Araujo et al. 2018] é a única base de dados em português com essa proposta, até então. Dessa forma, as principais contribuições deste trabalho são:

1. avaliação de três arquiteturas de rede neural, sendo estas a rede neural descrita por [Luz de Araujo et al. 2018], a biblioteca Flair [Akbik et al. 2019] e a biblioteca Spacy [Honnibal et al. 2020], que utiliza uma arquitetura otimizada para ambientes de produção. Além disso, no que diz respeito ao emprego da biblioteca Flair, foram utilizadas as representações (i.e., *embeddings*) IR introduzidas em [Akbik et al. 2018];

2. a produção de uma nova base de dados, baseada em petições iniciais, sendo a primeira base de dados deste tipo de documento jurídico anotada manualmente até o presente momento;

Este artigo está estruturado da seguinte forma: A seção 2 apresenta trabalhos relacionados ao problema de reconhecimento de entidade nomeada no domínio jurídico. A Seção 3 destaca a pergunta de pesquisa deste trabalho. Na seção 5, apresentamos a metodologia adotada para a realização dos experimentos. Em seguida, a seção 6 apresenta uma avaliação quantitativa dos resultados. Por fim, as conclusões e possíveis extensões a este trabalho são discutidas na seção 7.

2. Trabalhos relacionados

É possível classificar os principais modelos de reconhecimento de entidades nomeadas em quatro diferentes categorias [Yadav and Bethard 2019]: baseados em regra, baseados em probabilidade, baseados em aprendizado de máquina e híbridos. Modelos baseados em redes neurais têm se destacado bastante nos últimos anos, dado a sua maior capacidade de generalização, frente ao esforço de engenharia de conhecimento envolvido na produção de modelos baseados em regras. Esse tipo de modelo tem desempenho superior à modelos de *feature engineering* [Yadav and Bethard 2019]. Sistemas híbridos, por sua vez, que combinam tanto técnicas de *deep learning* quanto de probabilidade, como comprovado por [Lample et al. 2016], apresentam *scores* melhores que modelos baseados apenas em redes neurais.

A base de dados LeNER-br [Luz de Araujo et al. 2018], apresenta um corpus com anotações referentes a entidades do tipo pessoa, tempo, organização, jurisprudência, legislação e localização. Os autores demonstraram que o desempenho de um modelo BiLSTM-CRF treinado no seu conjunto de dados conseguia obter resultados competitivos com o estado da arte.

Em seu trabalho, [Son et al. 2016], utilizaram quatro modelos de redes neurais recorrentes (RNNs) na tarefa de reconhecimento de partes lógicas em documentos legais vietnamitas. Os autores ainda compararam a performance de três modelos baseados no algoritmo probabilístico *Conditional Random Fields* (CRF), quando combinados com modelos de RN, entre as quais se incluem redes neurais BI-LSTM-CRF, BI-LSTM e LSTM. Mesmo não usando *feature engineering*, as abordagens que se baseavam em redes neurais recorrentes conseguiram resultados competitivos com abordagens do estado-da-arte. O trabalho também mostrou o métodos híbridos, BI-LSTM-CRF, obteve o maior F_1 score entre os modelos testados, dando indícios da superioridade do método em tarefas de NER.

[Chen et al. 2020], por sua vez, propõem um modelo denominado BiLSTM-CRF-JCT para reconhecimento de entidades nomeadas em textos de processos judiciais. Os autores rotularam sentenças contendo entidades nomeadas chave desses textos e, então, converteram essas sentenças em vetores de palavras para, em seguida, usá-las como dados de entrada para uma rede neural BI-LSTM. Posteriormente, os resultados entregues pela rede BI-LSTM foram utilizados no treinamento de um modelo CRF, o qual ficou responsável por rotular, extrair e reconhecer entidades nomeadas em processos judiciais. Comparando com modelos relacionados, [Chen et al. 2020] mostra que a abordagem proposta se sobressai, alcançando um F_1 score que é aproximadamente 16% maior em contraste a seus pares.

Em [Cardellino et al. 2017], os autores propõem aprimorar a extração de informações em textos jurídicos mapeando termos do *LKIF-Core Ontology* na base de dados YAGO [Storks et al. 2019] e anotaram automaticamente exemplos com menções de entidades jurídicas. Foram treinados modelos de *support vector machine* (SVM), de redes neurais com e sem *word embeddings* e um outro utilizando o *Stanford NER* [Costa et al. 2017]. Os algoritmos baseados em *deep learning* mostraram desempenho superior quando comparados a algoritmos clássicos baseados em *feature engineering*.

Entretanto, estudos voltados a avaliação de modelos de NER em textos escritos em português ainda são bastante limitados. Além disso, até onde temos conhecimento, não existem trabalhos que avaliem arquiteturas de NER disponíveis em ferramentas voltadas a aplicações industriais, com foco em equilibrar acurácia e custo computacional, em textos jurídicos em português. Dessa forma, este trabalho avalia o desempenho de três arquiteturas de redes neurais distintas, a primeira replicada de acordo com [Luz de Araujo et al. 2018], a segunda fornecida por [Akbik et al. 2019] e a terceira, e última, disponibilizado pelo Spacy [Honnibal et al. 2020], sendo as duas últimas ferramentas populares no contexto de NER.

3. Pergunta de Pesquisa

Este artigo investiga o desempenho de três implementações de redes neurais para o problema de reconhecimento de entidade nomeada aplicado ao domínio jurídico.

PERGUNTA DE PESQUISA: *Das arquiteturas escolhidas, qual possui a melhor performance em termos de precisão, recall e F_1 score na tarefa de reconhecimento de entidades nomeadas no domínio jurídico?*

4. Métodos

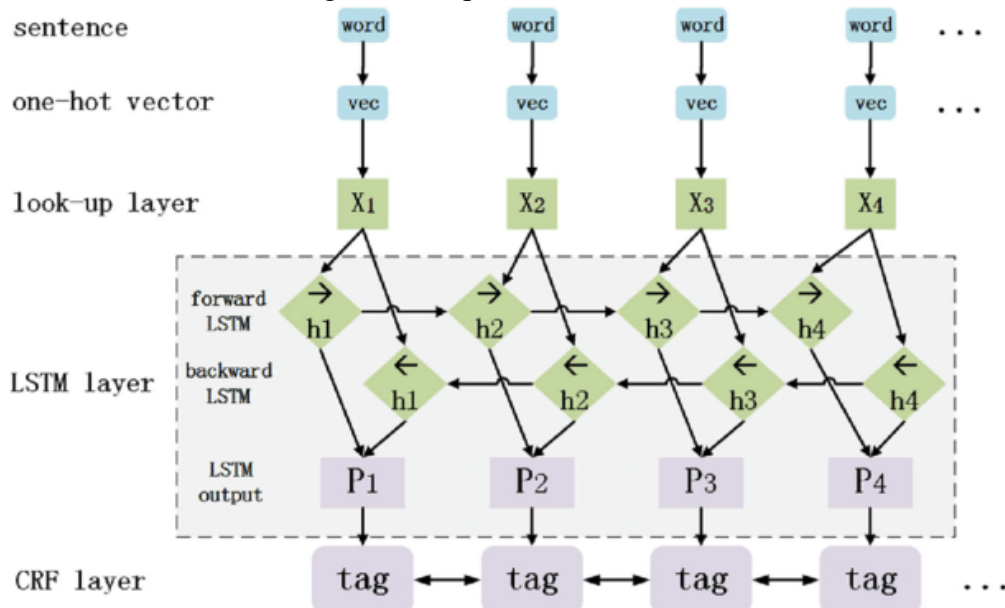
Nessa seção apresentamos as arquiteturas empregadas nesta pesquisa para treinar os modelos de reconhecimento de entidades nomeadas.

4.1. BiLSTM-CRF

Essa arquitetura combina uma rede neural *bidirectional long-short term memory* com um modelo baseado em probabilidade, o *conditional random fields* [Lample et al. 2016]. Ela foi usada em [Luz de Araujo et al. 2018] para avaliar o desempenho do *dataset* proposto e, também, é arquitetura padrão empregada pelo FLAIR. A proposta dele é utilizar um algoritmo que leve em conta uma quantidade virtualmente infinita de contexto e passar seus resultados para uma camada CRF, que por sua vez, se encarregará do rotulação das sequências de palavras, capturando as dependências ao longo dos rótulos.

Segundo [Panchendrarajan and Amaresan 2018], CRF mostrou-se ser bastante efetivo quando combinado com arquiteturas de redes neurais para tarefas de rotulação de sequências. Por outro lado, como ficou demonstrado em [Lample et al. 2016], arquiteturas híbridas, como utilizada para esta pesquisa, possui desempenho superior às demais. A Figura 1 mostra mais detalhadamente as camadas que compõem tal arquitetura.

Figura 1. Arquitetura BiLSTM-CRF

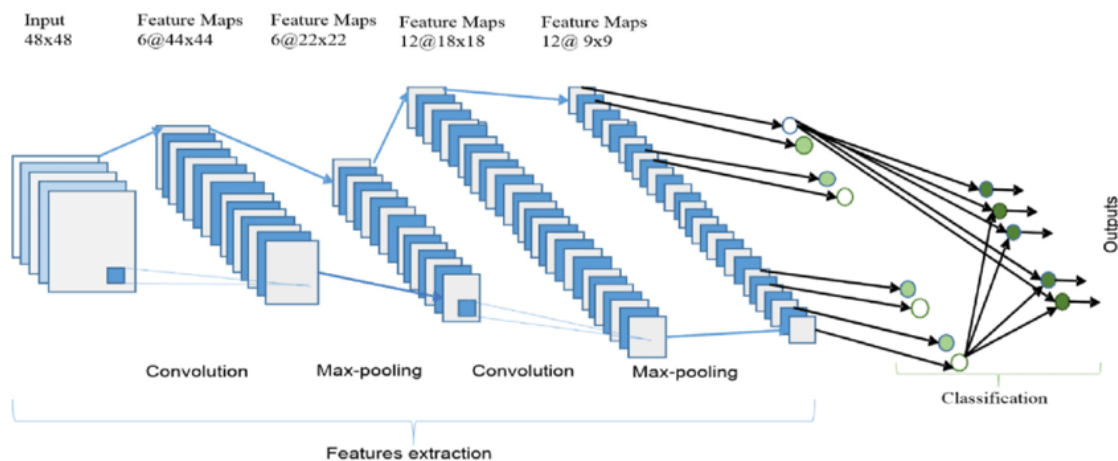


Fonte: [Ji et al. 2019]

4.2. Convolutional Neural Networks (CNN)

O Spacy utiliza como arquitetura interna para treinamento de modelos de NER uma CNN. De acordo com [Panchendrarajan and Amareesan 2018], CNNs foram aplicadas largamente no problema de modelagem de informações à nível de caractere, entre outras tarefas de processamento de linguagens naturais. A combinação de Bi-LSTM, CNN e CRF mostrou-se ser muito bem sucedida ao longo dos anos. Segundo [Yin et al. 2017], arquiteturas CNN são supostamente boas em extrair atributos de posição invariante, ao contrário de arquiteturas RNN que, por sua vez, que são melhores em modelo unidades em seqüências. Duas tarefas que se complementam em uma tarefa de rotulação de seqüências de palavras. Na Figura 2, é possível notar a estrutura tradicional de uma CNN.

Figura 2. Arquitetura CNN



Fonte: [Alom et al. 2019]

5. Metodologia Experimental

Neste trabalho é feita uma investigação acerca do uso de modelos híbridos e baseados em redes neurais para o problema de reconhecimento de entidades nomeadas no domínio jurídico. Para isso foram selecionadas três arquiteturas distintas fornecidas por [Akbik et al. 2019], [Luz de Araujo et al. 2018] e pelo *Spacy*. Os dois primeiros tratam-se de bibliotecas para processamento de linguagens naturais de propostas opostas. Enquanto o *FLAIR* tem o foco em facilitar a execução de experimentos para trabalhos científicos por pesquisadores, o *Spacy* se concentra na aplicação industrial de técnicas PLN. Neste artigo, foram treinados, avaliados e comparados três redes neurais: duas BiLSTM-CRFs com configurações para *word embeddings* diferentes e uma CNN com conexões residuais, utilizando uma base de dados própria e a base de dados construída em [Luz de Araujo et al. 2018].

5.1. Base de Dados

A base de dados foi extraída de 74 petições iniciais de processos que tramitaram na justiça brasileira, de acesso acesso privado, fazendo um total de 1676 trechos extraídos. As páginas das petições iniciais foram submetidas à um processo de OCR (*optical character recognition*) para extração do conteúdo textual que viria ser utilizado na composição do conjunto de treinamento empregado neste trabalho. Em contraste, a base de dados do LeNER-BR conta com um número um pouco menor de documentos, porém maior em relação à sentenças anotadas. Ademais, a base de dados de [Luz de Araujo et al. 2018] é menos específico quanto ao tipo de texto jurídico que o compõe, sendo o apresentando neste trabalho composto apenas por petições iniciais.

Tabela 1. Características da base de dados.

Propriedade	Proposto	LeNER-BR
Documentos	74	70
Classes	2	6
Trechos	1.676	10.392
Entidades anotadas	10352	N/D

A Tabela 1 apresenta outras informações sobre a base de dados. A base de dados proposta suporta dois tipo de classes: Legislação e Jurisprudência. Em contraste com o LeNER-BR, que possui mais classes, nesta pesquisa optou-se por dar foco às duas classes, entre as seis do LeNER-BR, que são mais relevantes para petições iniciais. Percebe-se, também, que, no total, para ambas as classes, existem 10352 anotações para cada uma delas. A Tabela 2 mostra a distribuição das entidades Legislação e Jurisprudência no conjunto de treino e de teste. É possível notar por meio dessas figuras que a base de dados encontra-se desbalanceada, o que torna o F_1 score uma métrica bastante confiável para medir o desempenho dos modelos a serem treinados.

Tabela 2. Distribuição das entidades nos conjuntos de treino e teste

Conjunto	Jurisprudência	Legislação
Treino	9011	1341
Teste	3398	427

Figura 3. Exemplo de sentenças anotadas

que causaram evidente transtorno e evidencia a má prestação de serviço , com violação ao princípio da transparência , da confiança e da boa-fé objetiva
insertos nos artigos 4º e 6º do CDC LEGISLACAO . Por todo o acima exposto , na forma do artigo 269 , I do Código de Processo Civil LEGISLACAO ,
conhecido e apelação não provida . (Apelação Cível 2009 01 1 075609-5 JURISPRUDENCIA APC Relator Desembargador JAIR SOARE .) Em relação ao
CONTRATO BANCÁRIO INVERSÃO DO ÔNUS DA PROVA CDC Possibilidade da inversão do ônus da prova com base no artigo 6º , VIII , do CDC
LEGISLACAO Reconhecido que o cliente tem direito de postular a exibição de documentos - Possibilidade de determinação pelo juiz incidentalmente no

5.2. Medidas de Avaliação

O conjunto de dados foi dividido em dois subconjuntos, um para treinamento, composto por 70% das amostras e outro para teste, composto por 30%, ambas selecionadas aleatoriamente. Para cada uma das arquiteturas utilizadas foram empregadas as configurações padrões oferecidas pelas bibliotecas. No caso da rede neural LSTM-CRF, foram utilizadas as especificações trazidas em [Luz de Araujo et al. 2018]. Como medidas de avaliação foram usadas as métricas de precisão, *recall* e *F₁ score*, descritas a seguir:

$$rec = \frac{TP}{TP + FN} \quad prec = \frac{TP}{TP + FP} \quad F_1 = 2 \cdot \frac{prec \cdot rec}{prec + rec},$$

onde TP, TF, FP e FN significam verdadeiro-positivo, verdadeiro-negativo, falso-positivo e falso-negativo, respectivamente. A métrica *recall* passa para o cientista uma ideia da parcela de exemplos positivos que foram recuperados pelos modelo. Por outro lado, a precisão comunica percentualmente qual a quantidade dos exemplos classificados como positivos são de fato positivos. O *F₁ score*, por fim, mostra uma relação harmônica entre precisão e *recall* para que possa-se reconhecer qual modelo possui o melhor balanceamento entre essas duas métricas.

5.3. Recursos de Hardware e Software

Os experimentos foram realizados em um computador com processador Ryzen 5 2600x, 16GB ram DDR4. O sistema operacional utilizado foi o Ubuntu 18.04. As bibliotecas utilizadas foram, Tensorflow versão 1.15.0, Keras versão 2.3.1, Numpy versão 1.18.3, Pandas versão 1.0.3, Fasttext versão 0.9.1.

6. Resultados

Os experimentos realizados considerando os dois conjuntos de dados e as três arquiteturas foram analisados em relação às medidas de avaliação indicadas na Seção 5. A Tabela 3 mostra as pontuações obtidas para as métricas de precisão, *recall* e *F₁ score* para cada um dos modelos utilizados.

Como pode ser observado, o modelo *Spacy* utiliza uma CNN (*Convolutional Neural Network* com conexões residuais e usa *Bloom Embeddings* para representar palavras. O FLAIR, por sua vez, usa uma BiLSTM-CRF (Bidirectional Long-Short Term Memory-Conditional Random Fields), que é o modelo que domina o estado-da-arte para tarefas de *sequence labeling*, segundo [Akbik et al. 2018]. As configurações utilizadas no treinamento do modelo do FLAIR seguiram as recomendações de configurações do modelo de maior desempenho apresentado por esse mesmo trabalho: uma combinação de BiLSTM com uma CRF layer e FLAIR embeddings, *forward* e *backward*, para geração dos vetores de palavras.

Por outro lado, para reproduzir o modelo apresentado por [Luz de Araujo et al. 2018], empregou-se o uso do *TensorFlow* [Abadi et al. 2015], uma biblioteca para aprendizagem de máquina publicada e distribuída pelo Google. A arquitetura utilizada consiste no mesmo modelo proposto por [Lample et al. 2016], o qual consiste e uma BiLSTM combinada com um algoritmo probabilístico, o CRF, e incorporações de palavras (i.e., *embeddings*) à nível de caractere.

Tabela 3. Acurácia, precisão, *recall* e F_1 score dos modelos para cada entidade.

Framework	Modelo	Entidade	Precisão	Recall	F_1 score
Spacy	CNN	Legislação	0.79	0.44	0.56
		Jurisprudência	0.78	0.11	0.20
<i>TensorFlow</i>	BiLSTM-CRF	Legislação	0.91	0.87	0.89
		Jurisprudência	0.74	0.44	0.55
<i>FLAIR</i>	BiLSTM-CRF	Legislação	0.93	0.85	0.89
		Jurisprudência	0.78	0.53	0.63

Pode-se observar que, de acordo com os resultados apresentados na Tabela 3 o modelo BiLSTM-CRF com *FLAIR embeddings* treinado com auxílio da biblioteca para PLN *FLAIR* apresentou resultados superiores com relação à precisão e F_1 score em ambas as categorias de entidades anotadas na base de dados analisada. Entretanto, o melhor score para *recall* não foi alcançado por ele, mas sim pelo modelo BiLSTM-CRF usado por [Luz de Araujo et al. 2018] em seus próprios experimentos. Contudo, vale ressaltar que a diferença entre as pontuações adquiridas pelo modelo do *FLAIR* e do *TensorFlow* não foram tão distantes uma da outra, o que, como pode ser visto pelo seu F_1 score, torna o modelo BiLSTM-CRF com *FLAIR embeddings* o modelo com melhor equilíbrio entre precisão e *recall*. O modelo CNN com conexões residuais, por sua vez, mostrou-se ter a pior performance entre seus pares. Quando comparado com o modelo treinado pelo *FLAIR*, ele não conseguiu se sobressair em nenhuma métrica. É importante notar, inclusive, que suas pontuações para *recall* e F_1 score foram bem menores que suas contrapartes. Comparando-o com o modelo de [Luz de Araujo et al. 2018], contudo, para a entidade de jurisprudência, o modelo do *Spacy* obteve uma precisão levemente superior ao modelo do *TensorFlow*.

Os resultados aqui expostos demonstram que abordagem que empregam modelos híbridos, ou seja, que combinam métodos probabilísticos e de aprendizado de máquina, conseguem trazer desempenhos superiores às que fazem uso de apenas de redes neurais. Além disso, eles mostram que o uso de *FLAIR embeddings* na geração dos vetores de palavras, que serão entregues como *input* para uma rede neural, pode prover uma melhora substancial nos scores do modelo a ser treinado. Os resultados, também, reafirmam a superioridade de modelos BiLSTM-CRF na tarefa de *sequence labeling* já demonstrada em artigos anteriores.

7. Conclusão

O presente trabalho tem como proposta um estudo sobre a aplicação de modelos baseados exclusivamente em redes neurais e modelos híbridos para o problema de reconhecimento de entidades nomeadas em textos legais. Assim sendo, foram avaliados três arquiteturas do estado da arte em tarefas de NER. O primeiro trata-se de uma CNN

com conexões residuais pré-implementada e distribuída pela biblioteca de PLN *Spacy*. O segundo é uma BiLSTM-CRF com *FLAIR embeddings* que foi treinada utilizando o *FLAIR* e, por fim, outro modelo BiLSTM-CRF implementado usando o *TensorFlow* [Luz de Araujo et al. 2018]. Esses modelos foram avaliados na base de dados proposta por [Luz de Araujo et al. 2018] e em uma base de dados própria, construída para este trabalho à partir de petições iniciais, composta por 74 documentos, perfazendo, ao todo, 1676 amostras.

Os resultados nas bases de dados analisadas, apontaram uma superioridade dos modelos híbridos em tarefas de NER, mais especificamente, arquitetura BiLSTM auxiliada por um modelo CRF para classificação de sequências de palavras. Além disso, com o respaldo do estado-da-arte, pode-se supor que o uso de *embeddings* contribuiria para uma melhor rotulação dessas sequências.

Por outro lado, o modelo com pior performance foi o CNN com conexões residuais, treinado no *Spacy*, que, quando comparado ao modelo de melhor desempenho, mostrou-se ser inferior em todos os aspectos considerados. De certa forma este resultado corrobora a percepção de que a biblioteca *Spacy* e a arquitetura CNN tem por característica a produção de modelos compactos, com baixos custos computacionais e de memória, durante as fases de treinamento e inferência [Honnibal et al. 2020]. Como trabalhos futuros, acredita-se que a expansão da base de dados traria melhores resultados no que se refere à tarefa de rotulagem de sequências de palavras, visto que a dinâmica da performance de redes neurais costuma flutuar para melhor ou pior em uma relação que é diretamente proporcional à quantidade de dados. Além disso, a ampliação do número de amostras pode solucionar o problema de desbalanceamento das classes Legislação e Jurisprudência na base de dados proposta, o que pode afetar direta ou indiretamente o desempenho dos modelos de NER a serem treinados. Ademais sugere-se o aumento ou até especialização do número de classes por meio da inclusão de entidades *fine-grained*.

Referências

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., and Vollgraf, R. (2019). Flair: An easy-to-use framework for state-of-the-art nlp. In *NAACL 2019, 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 54–59.
- Akbik, A., Blythe, D., and Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. In *COLING 2018, 27th International Conference on Computational Linguistics*, pages 1638–1649.
- Alom, M. Z., Taha, T., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M., Hasan, M., Essen, B., Awwal, A., and Asari, V. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8:292.

- Cardellino, C., Alemany, L. A., Teruel, M., and Villata, S. (2017). A low-cost, high-coverage legal named entity recognizer, classifier and linker. *Proceedings of the International Conference on Artificial Intelligence and Law*, pages 9–18.
- Chen, J., Huang, Y., Yang, F., and Li, C. (2020). A novel named entity recognition approach of judicial case texts based on BiLSTM-CRF. *12th International Conference on Advanced Computational Intelligence, ICACI 2020*, pages 263–268.
- Costa, C. M., Veiga, G., Sousa, A., and Nunes, S. (2017). Evaluation of stanford ner for extraction of assembly information from instruction manuals. In *2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 302–309. IEEE.
- Dale, R. (2019). Law and word order: Nlp in legal tech. *Natural Language Engineering*, 25(1):211–217.
- D’Angelo (2019). Em dois anos, número de startups jurídicas cresce 300% no brasil. Disponível em: <https://www.channel360.com.br/em-dois-anos-numero-de-startups-juridicas-cresce-300-no-brasil/>. Acesso em: 30 maio 2021.
- Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.
- Ji, B., Liu, R., Li, S., Yu, J., Wu, Q., Tan, Y., and Wu, J. (2019). A hybrid approach for named entity recognition in chinese electronic medical record. *BMC Medical Informatics and Decision Making*, 19.
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, pages 260–270.
- Luz de Araujo, P. H., de Campos, T. E., de Oliveira, R. R. R., Stauffer, M., Couto, S., and Bermejo, P. (2018). LeNER-Br: a dataset for named entity recognition in Brazilian legal text. In *International Conference on the Computational Processing of Portuguese (PROPOR)*, Lecture Notes on Computer Science (LNCS), pages 313–323, Canela, RS, Brazil. Springer.
- Minkov, E., Wang, R. C., and Cohen, W. (2005). Extracting personal names from email: Applying named entity recognition to informal text. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 443–450.
- Panchendrarajan, R. and Amaresan, A. (2018). Bidirectional lstm-crf for named entity recognition. In *PACLIC*.
- Son, N. T., Nguyen, L. M., Quoc, H. B., and Shimazu, A. (2016). Recognizing logical parts in legal texts using neural architectures. *Proceedings - 2016 8th International Conference on Knowledge and Systems Engineering, KSE 2016*, pages 252–257.
- Storks, S., Gao, Q., and Chai, J. Y. (2019). Recent advances in natural language inference: A survey of benchmarks, resources, and approaches. *arXiv preprint arXiv:1904.01172*.
- Yadav, V. and Bethard, S. (2019). A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. *arXiv*.

Yin, W., Kann, K., Yu, M., and Schütze, H. (2017). Comparative study of cnn and rnn for natural language processing. *arXiv preprint arXiv:1702.01923*.