

Classificação Automática de Textos Jurídicos

Orientador: Jomi Fred Hübner (Departamento de Automação)

Co-orientador: Aires José Rover (Departamento de Direito)

Contexto e Motivação

De acordo com o último relatório “Justiça em Números”¹, produção estatística publicada anualmente pelo Conselho Nacional de Justiça (CNJ)², o Poder Judiciário brasileiro finalizou o ano de 2017 com 80,1 milhões de processos em tramitação, isto é, aguardando solução definitiva. Em análise aos anos anteriores, houve um crescimento acumulado no período 2009-2017 de 31,9%, que equivale a um acréscimo de 19,4 milhões de processos.

À vista da litigiosidade crescente, o Poder Judiciário brasileiro, representado pelo CNJ, teve como iniciativa recente a criação do Centro de Inteligência Artificial (2019), em fase de implementação, com o objetivo de atrair esforços para melhorar a tramitação de processos e as decisões judiciais, além de combater a morosidade dessas tarefas.

Atualmente os processos judiciais são classificados e decididos manualmente pelo julgador e sua equipe. A tarefa de classificação, se desempenhada computacionalmente, poderá atingir resultados como a identificação de vieses e a divergência de decisões, além da própria rapidez. Essa classificação também poderá permitir a análise preditiva do resultado de processos judiciais futuros.

A dificuldade, porém, reside no tipo dos dados submetidos à classificação. Consistem em dados textuais não estruturados que, além de serem representados em língua natural, possuem vocabulário próprio da área jurídica.

¹ BRASIL. Conselho Nacional de Justiça (CNJ). **Justiça em Números 2018**: ano-base 2017. Brasília: CNJ, 2018. Disponível em: <http://www.cnj.jus.br/files/conteudo/arquivo/2018/08/44b7368ec6f888b383f6c3de40c32167.pdf>. Acesso em: 23 maio 2019.

² BRASIL. Conselho Nacional de Justiça (CNJ). Portaria n. 25 de 19 de fevereiro de 2019. Institui o Laboratório de Inovação para o Processo Judicial em meio Eletrônico – Inova PJe e o Centro de Inteligência Artificial aplicada ao PJe e dá outras providências. **Diário da Justiça eletrônico CNJ n. 35/2019**, Brasília, DF, 22 fev. 2019, p. 4-7. Disponível em: <http://www.cnj.jus.br/atos-normativos?documento=2829>. Acesso em: 23 maio 2019.

Objetivo

Investigar técnicas de aprendizado de máquina para classificar processos judiciais em três grupos: extinção, improcedente (rejeição do pedido), procedente (acolhimento do pedido). No caso procedente, o sistema deve prever o valor da condenação (considerando dano moral e material).

Escopo: processos relacionados ao consumo do serviço de transporte aéreo, para os quais temos aproximadamente 2000 processos entre os anos de 2015 a 2019. A base de dados é a do Juizado Especial Cível (Fórum) da UFSC.

Método

Dentre várias técnicas para resolver o problema proposto, este trabalho irá investigar diferentes técnicas de aprendizado de máquina para a classificação dos processos e previsão de valores de condenação (regressão). Duas abordagens serão consideradas (cf. descrição que segue) e comparadas.

O trabalho será realizado em conjunto com o grupo de pesquisa "Governo Eletrônico, Inclusão Digital e Sociedade do Conhecimento", do Programa de Pós-Graduação em Direito (PPGD/UFSC).

Abordagem 1

Um especialista identifica atributos relevantes para a classificação (gênero, idade, tempo de atraso do voo, cancelamento do voo, extravio de bagagem, condições climáticas, compromisso perdido, ...). Para cada processo exemplo, o especialista associa valores a esses atributos. O classificador tem os valores desses atributos como entrada.

A implementação dessa abordagem passa por duas fases:

- Extrair automaticamente os atributos de processos
- Classificar os processos
- Prever o valor da condenação

Abordagem 2

Uma segunda abordagem a testar é ignorar a etapa de identificação de atributos e sua valoração. A classificação do processo é baseada apenas no texto de entrada.

A implementação dessa abordagem passa por duas fases:

- Classificar os processos com base no texto
- Prever o valor da condenação