

Renato Melón: Calibragem da IA de linguagem nas petições

19/09/2023

A linguagem jurídica, com toda a sua gravidade e nuance, oferece um desafio único ao domínio da aprendizagem de máquinas. No cerne de cada petição legal está a convergência do pensamento humano intrincado e a representação linguística precisa.



O mundo jurídico exige não apenas clareza e precisão,

mas também persuasão e credibilidade. Na busca para estreitar o *gap* entre o pensamento humano e o conteúdo gerado por máquinas, o algoritmo de Otimização de Política Proximal (PPO) surge como uma ferramenta poderosa na afinação de Modelos de Linguagem de Grande Escala (LLMs) para esforços jurídicos. Afinal, o que é PPO?

A Otimização de Política Proximal (PPO) representa um avanço pioneiro no âmbito da aprendizagem por reforço, visando aprimorar como os modelos interagem e aprendem com seus ambientes. Como delineado por Schulman et al. em seu trabalho seminal "*Proximal Policy Optimization Algorithms*", o PPO introduz uma abordagem inovadora para os métodos de gradiente de política. Em vez da atualização convencional de um gradiente por amostra de dados, frequentemente vista nas técnicas padrão de gradiente de política, o PPO utiliza de forma inovadora uma função objetiva "substituta" para permitir múltiplas épocas de atualizações em minibatches.

Esta metodologia não apenas simplifica a implementação, mas também realça a complexidade da amostra, oferecendo um equilíbrio eficaz entre a intrincadeza do modelo e a execução no mundo real. Os testes extensivos de Schulman e sua equipe, abrangendo desde a locomoção robótica simulada até jogar jogos Atari, demonstram o desempenho superior do PPO em comparação com outras metodologias de gradiente de política online.

A evolução do pensamento para a linguagem

"A linguagem é o traje do pensamento." - Samuel Johnson.

Os seres humanos possuem uma habilidade notável de criar narrativas, uma habilidade essencial no campo jurídico. Todo pensamento germina como uma abstração intangível. Conforme amadurece, vestimo-lo com o tecido da linguagem, acrescentando-lhe contexto, emoção e intenção. Em petições legais, essa transformação é primordial. Não se trata apenas de declarar fatos; trata-se de apresentá-los de forma convincente. A jornada da cognição humana para expressão articulada é uma dança entre a concepção abstrata e a representação linguística.

Desde nossos primeiros grunhidos e gestos até as estruturas complexas das linguagens modernas, nossa capacidade de encapsular pensamentos em palavras tem sido central para nossa evolução como espécie. É o núcleo de como nos relacionamos com o mundo e uns com os outros. Filósofos há muito são fascinados por essa metamorfose.

Entre eles, Ludwig Wittgenstein se destaca com suas percepções transformadoras sobre a relação entre linguagem e pensamento. Em sua obra-prima, o *Tractatus Logico-Philosophicus*, Wittgenstein afirmou famosamente: "*Os limites da minha linguagem significam os limites do meu mundo*". Isso sublinha a ideia de que nossa linguagem não apenas reflete nossa realidade, mas em muitos aspectos, a molda. Nossa percepção do mundo, nosso próprio entendimento da existência, é vinculado e moldado pelos constructos linguísticos que empregamos.

No contexto da arena jurídica, esta base filosófica torna-se ainda mais comovente. Quando advogados redigem petições, eles não estão apenas construindo frases; estão moldando percepções, moldando a realidade usando o barro da linguagem. A escolha de cada palavra e frase não é apenas uma decisão linguística, mas também filosófica, ecoando a ideia de Wittgenstein da linguagem como um espelho de nossa visão de mundo.

Além disso, os trabalhos posteriores de Wittgenstein, *Investigação Filosófica*, exploraram o conceito de "jogos de linguagem", sugerindo que nosso uso da linguagem está profundamente entrelaçado com as atividades de nossa vida e, assim, cada profissão, incluindo a lei, tem seu próprio conjunto de regras e nuances em suas expressões linguísticas. Neste cenário, a transformação do pensamento em linguagem não é apenas um exercício cognitivo. É um balé intricado de cognição, linguística e filosofia, com cada disciplina lançando luz sobre a profundidade da expressão humana. Portanto, quanto às petições legais: Considere a jornada de um advogado ao redigir uma petição. Eles começam com um argumento central, enriquecendo-o constantemente com estatutos, jurisprudências e evidências. A escolha de cada palavra e frase é deliberada e proposital. Esta arte linguística é semelhante à composição de um músico ou à obra-prima de um pintor.

PPO e LLM: compondo a sinfonia jurídica

"Tudo deve ser feito o mais simples possível, mas não mais simples que isso." - Albert Einstein.

"PPO é como quando você tem uma ideia aproximada do que está tentando expressar, mas ainda não sabe quais palavras escolher." (Ehsan Kamalinejad)

Embora os LLMs como o GPT-4 da OpenAI demonstrem capacidades linguísticas impressionantes, a linguagem jurídica exige tratamento especializado. O *Proximal Policy Optimization* (PPO) atua como um maestro, direcionando as decisões do LLM para gerar



conteúdo jurídico coerente, contextualmente apropriado e persuasivo. A essência do PPO reside na otimização da política que um LLM usa para selecionar palavras ou tokens.

Assim como um mentor guiando um advogado novato, o PPO refina a trajetória do modelo, garantindo que suas escolhas linguísticas ressoem com as complexidades jurídicas, seja em geral ou em petições legais LLM.

Apesar das notáveis realizações do PPO no campo dos problemas de controle contínuo desde sua criação em 2017, várias críticas merecem atenção. Embora o algoritmo tenha mostrado seu poder em problemas de controle clássico e jogos Atari, tornando-se um padrão dentro de sua classificação de problemas de referência, sua trajetória não esteve isenta de obstáculos.

Van Heeswijk destaca o notável equilíbrio do PPO entre velocidade, cautela e usabilidade, enfatizando a inclinação do algoritmo para uma convergência rápida e aprimorada em comparação com seus pares. No entanto, por trás dessa fachada de simplicidade e eficiência, existem desafios inerentes. Ao contrário de seus equivalentes, como gradientes naturais e TRPO, o PPO carece de garantias teóricas sólidas. Tal ausência, embora aparentemente inconsequente diante de seu sucesso empírico, apresenta potenciais armadilhas em aplicações que exigem um rigoroso embasamento matemático.

Depender apenas de evidências empíricas, especialmente em cenários que requerem alta confiabilidade e previsibilidade, pode ser um empreendimento precário. Além disso, o adágio de que "*a simplicidade compensa no Deep Reinforcement Learning*", conforme insinuado pela natureza paradoxal do design do PPO, não está isento de limitações. Estruturas simplistas podem agilizar processos e reduzir sobrecargas computacionais, mas também podem ignorar detalhes e nuances intrincados inerentes a tarefas mais complexas.

Embora o PPO prospere em certos ambientes, sua aplicabilidade ampla permanece um tópico de controvérsia entre os pesquisadores. Situações complexas do mundo real, repletas de nuances e variáveis, podem desafiar a arquitetura do PPO, o que pode levar a decisões menos ótimas ou até mesmo errôneas. Adicionalmente, a rápida evolução da paisagem do aprendizado de máquina apresenta outro dilema.

À medida que surgem novos algoritmos, o uso de *benchmarks* apenas em problemas clássicos de controle e jogos Atari pode tornar o PPO obsoleto ou menos eficiente. Há uma necessidade imperativa de recalibrar e reavaliar constantemente a relevância do PPO diante dos desafios sempre em evolução. Embora o domínio e o sucesso do PPO em domínios específicos permaneçam inegáveis, um exame holístico revela certas limitações e desafios.

Como qualquer algoritmo, a jornada do PPO é uma interação contínua entre forças e limitações, exigindo constante introspecção e evolução. No reino específico das petições jurídicas, as nuances da linguagem, precedentes e contexto são de extrema importância. A essência de um argumento jurídico não reside apenas em sua veracidade factual, mas em sua apresentação, tom e persuasão. Dadas as críticas ao PPO, sua aplicação na elaboração de petições jurídicas merece um escrutínio mais detalhado. A ausência de garantias teóricas robustas no PPO, em contraste com algoritmos como gradientes naturais e TRPO, levanta preocupações. Petições jurídicas muitas vezes requerem fundamentações lógicas rigorosas, e qualquer inconsistência

ou negligência pode ter implicações profundas no caso em questão. Dependendo apenas do sucesso empírico do PPO sem um sólido embasamento teórico pode levar a possíveis deturpações ou interpretações errôneas em argumentos jurídicos. Além disso, a simplicidade do PPO, embora benéfica em certos domínios, pode não ser suficiente para capturar as multifacetadas intricacias da linguagem e raciocínio jurídicos.

O vasto tapete da lei é tecido com inúmeros fios de precedentes, estatutos e interpretações. Uma abordagem que serve para todos, como pode ser a armadilha de um algoritmo simplista, corre o risco de ignorar detalhes cruciais. Por exemplo, o PPO pode gerar um argumento genérico que perde nuances sutis, mas vitais, que um modelo mais complexo ou especializado poderia captar. Além disso, a paisagem jurídica, assim como o campo do aprendizado de máquina, está em constante evolução. Novas decisões, legislações e interpretações surgem, redesenhando os contornos da prática jurídica. Dependendo apenas dos benchmarks atuais do PPO pode tornar suas petições geradas desatualizadas ou até mesmo desalinhadas com o pensamento jurídico contemporâneo.

Em conclusão, enquanto as realizações do PPO em domínios específicos são louváveis, sua aplicação no campo matizado das petições jurídicas apresenta desafios. Isso destaca a necessidade de uma abordagem cuidadosa e bem calibrada, potencialmente combinando os pontos fortes do PPO com outros algoritmos ou supervisão humana. À medida que a tecnologia se entrelaça cada vez mais com a lei, garantir a precisão, relevância e profundidade do conteúdo jurídico gerado por máquinas permanece de suma importância.

A alquimia do PPO em petições jurídicas

Para visualizar o poder transformador do PPO em petições jurídicas, ilustremos com um exemplo.

Suponha que um LLM está produzindo uma petição referente a uma violação ambiental. Sem a otimização PPO, ele pode gerar:

"A Empresa X está causando poluição ao liberar resíduos no rio".

No entanto, uma vez refinado com o PPO adaptado para linguagem jurídica, ele pode afirmar:

"As ações da Empresa X, caracterizadas pela descarga indiscriminada de resíduos perigosos no rio, violam manifestamente as proteções ambientais delineadas em [regulamentação específica], justificando assim uma intervenção judicial imediata".

Observe a metamorfose? A frase enriquecida não é apenas mais clara; ela é estratificada com precisão, persuasão e referência legal. Usar a Otimização de Política Proximal (PPO) no contexto de elaboração de petições jurídicas implica em entender como sua formulação matemática subjacente pode ser aplicada ao universo matizado da linguagem e raciocínio jurídico.

Aqui está uma exploração mais detalhada: Em sua essência, o PPO busca maximizar uma certa função objetivo. No mundo da Aprendizagem por Reforço (RL, em inglês), isso frequentemente significa maximizar a recompensa cumulativa esperada ao longo do tempo. Para petições



jurídicas, essa "recompensa" pode ser conceituada como a eficácia de um argumento redigido, talvez medida em termos de sua persuasão, coerência ou alinhamento com precedentes jurídicos. Matematicamente, vamos representar este objetivo como:

$$J(\theta) = \mathbb{E} \pi_{\theta}[R(t)]$$

Onde:

θ é o conjunto de parâmetros que governa nosso modelo PPO.

π_{θ} representa a política (ou estratégia de tomada de decisão) derivada desses parâmetros.

$R(t)$ é a recompensa no tempo t , simbolizando a eficácia de uma determinada frase ou parágrafo em nossa petição.

A magia do PPO reside no seu uso de uma função objetivo substituta. Em vez de maximizar diretamente o objetivo principal, o PPO cria uma função substituta que oferece um equilíbrio entre exploração (testando novas estruturas linguísticas) e exploração (mantendo estratégias eficazes conhecidas). A função substituta pode ser representada como:

$$LCLIP(\theta) = \mathbb{E} \pi_{\theta}[\min(\pi_{\theta}^{old}(a|s) \pi_{\theta}(a|s) A_t, \text{clip}(\pi_{\theta}^{old}(a|s) \pi_{\theta}(a|s), 1-\epsilon, 1+\epsilon) A_t)]$$

Aqui: A_t é a função de vantagem, que representa o quanto uma ação a é melhor em comparação com a média no estado s .

A função de corte (clip) garante que as atualizações da política permaneçam dentro de um intervalo especificado, ditado por ϵ , evitando mudanças drásticas. No contexto de petições legais, "ações" podem ser a seleção de terminologias jurídicas específicas, argumentos ou referências a decisões anteriores. O "estado" representa o contexto atual ou o ponto que está sendo feito na petição. Uma política ideal escolheria palavras e estruturas que maximizam a persuasão e coerência da petição. Treinar o modelo PPO envolve iterar sobre esta função substituta, atualizando gradualmente os parâmetros θ usando ascensão de gradiente estocástico para aumentar a persuasão do texto legal gerado. Cada iteração usaria uma extensa base de dados de textos legais, precedentes e argumentos, refinando o entendimento e representação do modelo de linguagem jurídica eficaz.

Finalmente, enquanto as formulações matemáticas do PPO foram originalmente projetadas para desafios no domínio da Aprendizagem por Reforço, sua potencial aplicação a petições legais é um testemunho da versatilidade do algoritmo. Ao reinterpretar e adaptar essas construções matemáticas às peculiaridades da redação jurídica, embarcamos em uma jornada fascinante onde os reinos da matemática, direito e inteligência artificial convergem.

PPO: A Simbiose de Eficiência e Exploração

"Por três métodos podemos aprender sabedoria: Primeiro, pela reflexão, que é a mais nobre; Segundo, pela imitação, que é a mais fácil; e terceiro pela experiência, que é a mais amarga." - Confúcio.

A beleza do PPO reside em seu equilíbrio entre exploração (tentar coisas novas) e exploração (usando estratégias conhecidas). Isso reflete a forma como os profissionais do direito constroem argumentos. Às vezes, eles se baseiam em precedentes estabelecidos (exploração) enquanto, em outras vezes, elaboram interpretações novas (exploração). O PPO garante que o LLM não fique preso em uma rotina linguística, repetindo constantemente frases semelhantes. Em vez disso, incentiva o modelo a experimentar, expandindo os limites, mas permanecendo dentro dos confins do léxico jurídico.

Além da linguagem: as implicações éticas

"A verdadeira questão não é se as máquinas pensam, mas se os homens o fazem." - B.F. Skinner.

Em nossa incessante busca pelo avanço tecnológico, LLMs, aprimorados e afiados pelo PPO, agora navegam no delicado tecido da linguagem jurídica com precisão. Mas aqui reside uma junção crítica, um cruzamento de capacidade de máquina e consequência moral. Esses modelos, hábeis em elaborar petições e construir argumentos, ainda enfrentam uma questão central: um algoritmo pode verdadeiramente compreender a essência inerente da justiça?

A experiência humana é um mosaico de emoções, julgamentos e intenções. Os humanos, com todas as suas complexidades, podem oscilar imprevisivelmente entre benevolência e malevolência. Há indiscutíveis instâncias em que as ações humanas, manchadas por preconceitos e preconceitos, resultam em grave dano a outros. Um PPO bem ajustado, em sua pureza matemática, tem o potencial de transcender essas falhas humanas e evitar as armadilhas da turbulência emocional. No entanto, evitar a falibilidade humana significa alcançar a justiça?

É possível que, ao contornar as falhas humanas, corramos o risco de ignorar a profundidade da empatia humana, os nuances do julgamento moral e o entendimento quintessencial da justiça? Confiar apenas na tecnologia, não importa o quão avançada, levanta preocupações inquietantes. Não se trata apenas da capacidade de replicar articulações legais semelhantes às humanas, mas sobre a profunda responsabilidade e discernimento envolvidos no exercício da justiça. E, enquanto a tecnologia pode, sem dúvida, auxiliar e aumentar, a santidade do toque humano, sem sombra de dúvida, permanece suprema, especialmente em arenas tão santificadas quanto o direito.

Conclusão: confluência de homem e máquina

A chegada do PPO na otimização de LLMs para aplicações jurídicas sinaliza um momento empolgante onde a intelectualidade humana encontra a eficiência da máquina. Ao entender como os humanos transformam pensamentos em linguagem e replicar isso com PPO em LLMs, nos aproximamos de um futuro onde as máquinas não apenas entendem nossas palavras, mas também a profundidade por trás delas. No entanto, ao nos encontrarmos neste cruzamento, é imperativo proceder com cautela, valorizando a essência humana que dá alma ao direito — e talvez o torne cada vez mais necessário para se proteger as partes vulneráveis da sociedade.

Fonte: <https://conjur.jumps.com.br/2023-set-19/renato-melon-calibragem-ia-linguagem-peticoes/>