

DESAFIOS DE APRENDIZADO DE MÁQUINA PARA CIDADES INTELIGENTES

A ÁREA DE APRENDIZADO DE MÁQUINA APRESENTA MUITOS DESAFIOS, ESPECIALMENTE QUANDO OS DADOS VARIAM COM O TEMPO.

.....
por Junior Barrera, Marcelo Finger e Roberto Hirata Jr.
.....

Entre os sistemas de Computação mais importantes para as cidades inteligentes estão os sistemas que aprendem através de exemplos, isto é, que integram de alguma forma métodos de Aprendizado de Máquina (ML - *Machine Learning*). A popularização dos aparelhos celulares, dos sensores, dos dispositivos da Internet das Coisas (IoT) e também das câmeras tem trazido para o cenário da administração pública e privada, assim como para o cidadão comum, um conjunto de dados sem precedentes. As aplicações vão desde vigilância, num sentido bem amplo (pessoas, dados, geolocalização), até organização, prevenção de riscos, etc. Os desafios para a área de Computação são enormes e vão desde a definição de protocolos para a coleta e organização dos dados, o armazenamento de dados e análises, até os sistemas de tomada de decisão.

Os sistemas de decisão baseados em ML dependem de características numéricas e/ou qualitativas (alto/baixo, congestionado/livre, lotado/cheio/vazio, etc.) que são extraídas manualmente, auxiliados por especialistas ou automaticamente, baseado em métodos de seleção de características ou usando redes neurais profundas, principalmente as convolucionais quando os dados são imagens. Os objetos de estudo devem ser representados formalmente (por exemplo, como vetores em que, se n for o número de características, os objetos são representados por pontos no espaço R^n).

De posse dessa representação, diversas análises de dados são possíveis. Podem-se classificar objetos de uma amostra associada a um problema de interesse (onde um conjunto de objetos X é associado a um conjunto Y , também conhecido como rótulos de X), podem-se analisar os dados como agrupamentos, ou outras formas de visualização, ou pode-se aplicar um algoritmo de ML para induzir um classificador, baseado na hipótese de que existe uma distribuição de probabilidade conjunta entre X e Y ¹. Existem diversos algoritmos e estratégias para se induzir um classificador a partir de conjuntos de amostras (mais conhecidas por conjuntos de treinamento) de X e Y . Escolher um algoritmo, ou mesmo escolher as melhores características que favoreçam o sucesso de um classificador (conceito que envolve estabelecer um custo para cada classificação errada quando se aplica o classificador e cuja explicação detalhada foge ao escopo deste texto), é por si

Entre os sistemas de Computação mais importantes para as cidades inteligentes estão os que integram métodos de Aprendizado de Máquina.

só um desafio, principalmente quando estamos trabalhando com conjuntos de treinamento, ou com uma quantidade de rótulos muito grandes.

Um segundo desafio é lidar com dados escassos, com poucos dados rotulados, pois a rotulagem é uma tarefa onerosa normalmente feita por seres humanos. Nesse sentido, há uma tendência a se criarem padrões prata (em contraposição ao padrão ouro, que é a rotulação feita por seres humanos), ou usar algoritmos de ML que se ajustam de acordo com os erros, que é o caso dos algoritmos de aprendizado por reforço.

Um terceiro desafio ainda maior ocorre quando há uma variabilidade temporal associada ao conjunto X , ou ao conjunto Y , ou a ambos (como se X e Y agora fossem X^t e Y^t). Nesses casos, o classificador precisa se adaptar à variabilidade dos dados, empregando modelos dinâmicos, cuja aplicação é ainda pouco explorada na área de ML. Facilita-se a análise nesses casos se os dados apresentarem periodicidade, que é o caso em dados de transporte, de clima e temperatura, etc.

Finalmente, não podemos deixar de mencionar o desafio da integração colaborativa desses sistemas, realizando fusão de análises de dados. Como é sabido, encontrar soluções ótimas individuais não garante e, em geral, dificulta o problema de se achar soluções globais integradas. Nessa linha, as pesquisas ainda são muito restritas a uma área conhecida como *agentes inteligentes* e há muito espaço até se alcançar uma plena integração de sistemas de ML.

Não poderíamos terminar este artigo sem falar sobre os *desafios éticos envolvidos*, que consideramos ser um dos maiores que enfrentaremos em relação aos dados relacionados às cidades inteligentes e que, embora afetem todas as áreas da Computação envolvidas, os sistemas de ML podem ajudar. Os dados não sensíveis (que não têm uma identificação do ser humano) e sistemas de ML deveriam ser abertos, ou pelo menos auditáveis por seres humanos, ou melhor ainda, por outros sistemas de ML abertos. A tarefa desses sistemas seria evitar favorecimento ou desfavorecimento, econômico ou social, de determinados indivíduos, gêneros, etnias, etc. Os dados produzidos pelos cidadãos ou pelo governo pertencem à população e, se podem ser compartilhados sem ferir a privacidade do cidadão, devem ser compartilhados abertamente. Pesquisar formas de auditar automaticamente, ou fiscalizar desvios nas condutas éticas dos sistemas de ML, é, talvez, o principal desafio para o futuro da convivência da humanidade com os sistemas inteligentes. ●

Referências

1 Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009, Second edition.



JUNIOR BARRERA | É professor titular do Departamento de Ciência da Computação do Instituto de Matemática e Estatística da Universidade de São Paulo. Desenvolve pesquisas em Morfologia Matemática, Estimação de Operadores Morfológicos, Identificação e Modelagem de Sistemas Dinâmicos de Reticulado, Vídeo Digital, Modelagem de Sistemas Biológicos e Esportes Coletivos de Invasão.



MARCELO FINGER | É professor titular do Departamento de Ciência da Computação do Instituto de Matemática e Estatística da Universidade de São Paulo, bolsista (PQ) 1C do CNPq, editor do São Paulo Journal of Mathematical Sciences. Desenvolve pesquisas em Lógica, Inteligência Artificial, Bancos de Dados, Linguística Computacional e Humanidades Digitais.



ROBERTO HIRATA JR. | É professor associado do Departamento de Ciência da Computação do Instituto de Matemática e Estatística da Universidade de São Paulo. Desenvolve pesquisas em Morfologia Matemática, Processamento de Imagens e Vídeos e Aprendizado de Máquina.