Vinícius Carvalho Rosa

**Resenha da Monografia “Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy”**

O artigo intitulado *“Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy”* que em 2017 foi publicado na revista “***IEEE Intelligent Systems***” e desenvolvido por cinco autores acadêmicos ([Jan Kremer](https://ieeexplore.ieee.org/author/37086104276); [Kristoffer Stensbo-Smidt](https://ieeexplore.ieee.org/author/37086102090); [Fabian Gieseke](https://ieeexplore.ieee.org/author/37857080600); [Kim Steenstrup Pedersen](https://ieeexplore.ieee.org/author/37086108251); [Christian Igel](https://ieeexplore.ieee.org/author/37281518600)) afiliados à “University of Copenhagen” que está situada na Dinamarca, teve como objetivo discorrer sobre as ferramentas computacionais que auxiliam no desenvolvimento dos estudos da Astrofísica e Cosmologia, sobretudo os algoritmos de Aprendizado de Máquina *(Machine Leaning)* e análise de imagens que trabalham em função de grandes quantidades de dados *(Big-Data)* coletadosa partir das pesquisas observacionais dos telescópios mencionados, bem como: Very Large Telescope (VLT), Sloan Digital Sky Survey (SDSS), Visible and Infrared Telescope for Astronomy (VISTA), Large Synoptic Survey Telescope (LSST) e Thirty Meter Telescope (TMT).

A astronomia moderna é fundamentalmente precisa em termos observacionais quando comparado aos dados das pesquisas a duas décadas atrás, pois, as observações a fim de coletar dados, eram realizadas em uma grande escala de tempo que poderiam durar entre dias a meses, enquanto hoje, os dados podem ser adquiridos em apenas uma noite de observação. O telescópio Sloan Digital Sky Survey (SDSS), por exemplo, produz a cada noite ~200 GB de dados, que até o momento possibilitou obter quase um milhão de imagens do espaço. A partir de análises profundas dos dados obtidos pelo telescópio SDSS, foi possível detectar mais de 200 milhões de galáxias e mais as estrelas. Segundo os autores, o motivo pelo qual a precisão das pesquisas astronômicas é atualmente considerada promissoras, deve-se ao fato do avanço exponencial tecnológico nos últimos anos, sobretudo com a ascensão da tecnologia de aprendizado de máquina, que a partir de dados tendenciosos o algoritmo aprende a tratar e lidar com os rótulos e ruídos, que automatiza algumas das funções que antes eram realizadas de forma manual pelos pesquisadores.

Através da 1ª figura inserida no artigo, na qual representa o volume de dados existentes e futuros dados dos telescópios mencionados, evidenciam que as taxas de dados aumentaram e continuarão aumentar à medida que novas pesquisas são iniciadas. No entanto, segundo os autores, quanto mais dados obtidos das imagens que são produzidas a cada noite, mais eficiente e precisa devem ser as detecções dos fenômenos que ocorrem no universo, que para isso, é utilizado nas observações, uma técnica chamada; espectroscopia. A espetroscopia mede a contagem de fotos em milhares de comprimentos de onda. O espectro resultante permite identificar componentes químicos do objeto observado e, assim, permite determinar muitas propriedades interessantes.

Para analisar os dados obtidos em larga escala, é necessário utilizar os métodos de aprendizado de máquina, pois através da utilização dessa ferramenta tecnológica é possível descobrir a relação entre a entrada de dados (por exemplo, imagem de galáxias) e resultados (por exemplo, propriedades físicas de galáxias). Sendo assim, com base em amostras de entrada-saída, os pesquisadores relatam que esta técnica provou ser bem-sucedida em diversos contextos astronômicos. Por exemplo, as distâncias da terra às galáxias e estrelas, são medidas através do efeito Doopler, que se refere ao seus redshifts (deslocamento para o vermelho) e blueshifts (deslocamento para o azul). Se o objeto observado emitir na faixa da luz visível algum desvio para o vermelho, significa que o objeto está se distanciando da terra, logo está mais distante. Todavia, para relatar com eficiência as estimativas das distancias entre os corpos é necessário utilizar a técnica de espectroscopia, mas, em segundo plano, pode-se receber precisamente os redshifts utilizando a técnica de fotometria somente, que por sua vez é uma tarefa essencial, mas não resolvida, para os quais os métodos de aprendizado de máquina são aplicados. Segundo os autores, a técnica de fotometria que se utilizam métodos de aprendizado de máquina para medir distancias de objetos do cosmo, estão longe de estar no mesmo nível da espectroscopia. Sendo assim, algoritmos melhores e mais rápidos são muito desejados.

Por conseguinte, através do crescimento constante de grandes quantidades dados coletados em observações astronômicas, nos próximos anos, será essencial utilizar os métodos de aprendizagem de máquina para analisar imagens com alta precisão que processam terabytes de dados em tempo real!

**REFERÊNCIA**

J. Kremer, K. Stensbo-Smidt, F. Gieseke, K. S. Pedersen and C. Igel, "Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy,". **IEEE Intelligent Systems**, vol. 32, 2017. Disponível em:<https://ieeexplore.ieee.org/abstract/document/7887648 >.