

Identificando estilos de pinturas artísticas com entropia e complexidade de permutação

Adriano Henrique Danhoni Neves (PIBIC/CNPq), Alexandre de Ramos Alabora (PIBIC/CNPq), Haroldo Valentin Ribeiro (Orientador), e-mail: ra102561@uem.br.

Universidade Estadual de Maringá / Centro de Ciências Exatas/Maringá, PR.

Física Geral / Física Estatística e Termodinâmica.

Palavras-chave: Entropia, Complexidade, Arte.

Resumo:

Neste trabalho revisitamos o uso de medidas de entropia e complexidade de permutação aplicadas ao estudo de obras de arte. Analisando 8904 imagens de pinturas de quatro estilos artísticos (Barroco, Renascimento, Pontilhismo e *Hard-Edge Painting*), mostramos que medidas simples inspiradas em Física podem capturar aspectos qualitativos de obras de arte e produzir uma classificação de boa precisão. Em uma tarefa de regressão logística, verificamos que as medidas de entropia e complexidade distinguem arte barroca da renascentista com 76% de precisão e Pontilhismo de *Hard-Edge Painting* com 77% de acurácia.

Introdução

O estudo de arte e história da arte tem sido realizado predominantemente de maneira qualitativa e comparativa. Embora os resultados produzidos por essas investigações seja de enorme valia, esse tipo de abordagem limita muito a quantidade de obras que especialistas em pinturas podem lidar simultaneamente. Ao mesmo tempo, essa característica dos estudos de arte também impede uma visão mais global que seja baseada em evidências empíricas e em medidas quantitativas.

O acúmulo crescente de informações e dados em plataformas digitais sobre arte tornou possível empregar ferramentas quantitativas para caracterizar obras de arte. No caso particular de pinturas, existem algumas galerias virtuais que possibilitam acesso a grandes conjuntos de imagens de obras de arte. Nesse projeto de iniciação científica, revisitamos um estudo utilizando dados da enciclopédia *online* WikiArt.org para tentar quantificar obras de arte usando medidas de entropia e complexidade de permutação [1]. No trabalho, verificamos que essas duas medidas simples (extraídas diretamente das imagens) capturam informações relevantes das obras de arte; o que, por sua vez, permite uma classificação de estilos artísticos de forma automatizada usando métodos de aprendizagem estatística [2].

Materiais e Métodos

Dados

Os dados utilizados nesse estudo foram obtidos da página WikiArt.org de forma automatizada e compreendem 8904 imagens de obras de arte (pinturas) de quatro estilos: Barroco (6122 imagens), Renascimento (1616 imagens), Pontilhismo (550 imagens) e *Hard-Edge Painting* (616 imagens). Os arquivos foram obtidos no formato JPEG com 24 bits por pixel e, posteriormente, convertidos para escala de cinza usando a transformação de *luminance*. De modo efetivo, cada imagem pode ser pensada como uma matriz na qual cada elemento representa um pixel cujo valor (intensidade) pode variar entre zero e um.

Entropia e Complexidade de Permutação

As medidas de entropia e complexidade de permutação são estimadas a partir de uma distribuição de probabilidade associada a possíveis padrões de ordem em partições da imagem original [3]. A entropia de permutação H foi normalizada de modo que $H=0$ representa uma imagem muito ordenada com ocorrência de um único padrão ordinal. Por outro lado, $H=1$ representa uma imagem bastante irregular na qual todos os padrões ordinais ocorrem com igual probabilidade. Já a complexidade de permutação C mede o grau de "estrutura" da imagem, sendo que para um dado valor de entropia, a complexidade pode variar de um valor mínimo a um máximo, capturando informações complementares à entropia. Após processados, nossos dados compreendem os valores de entropia H , complexidade C e estilo (Barroco, Renascimento, Pontilhismo ou *Hard-Edge Painting*) de cada pintura. A Figura 1, mostra os valores de H e C de cada pintura, na qual as cores referem-se aos quatro diferentes estilos.

Tarefa de classificação logística

Utilizamos o método de classificação logística [3] para tentar prever o estilo de obras de arte em duas tarefas simples: a) distinguir entre arte barroca e renascentista e b) diferenciar pinturas do Pontilhismo de pinturas no estilo *Hard-Edge Painting*. Para isso, separamos 80% dos dados para treinar o classificador logístico e os 20% restantes foram usados para estimar a acurácia do modelo. Todos os modelos foram desenvolvidos usando o módulo de *machine learning scikit-learning* [4] da linguagem *Python*.

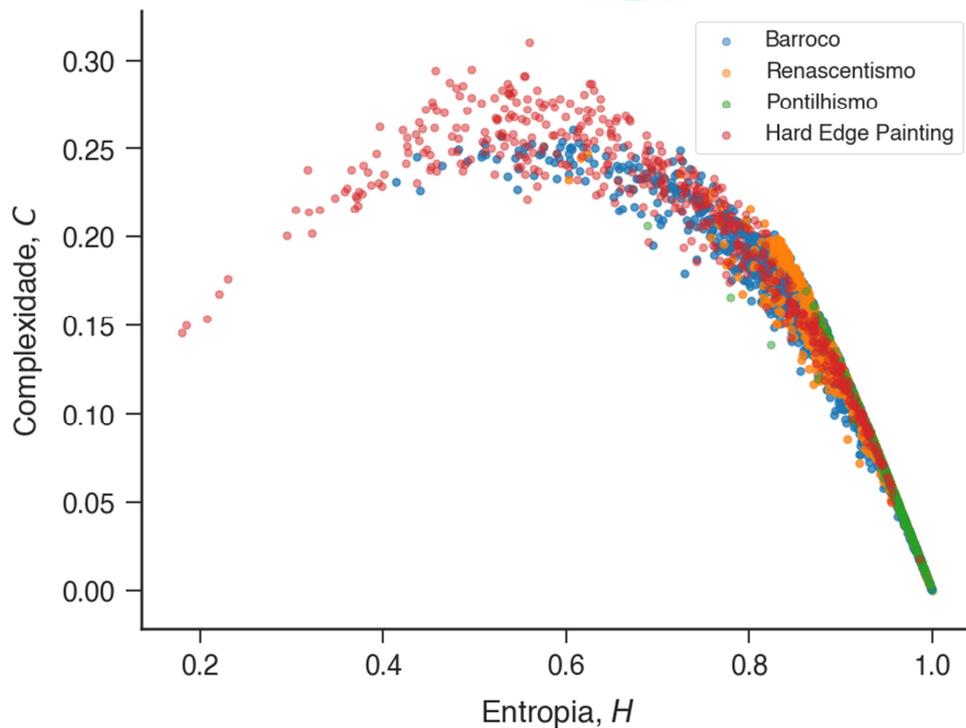


Figura 1 – Plano complexidade-entropia. Os pontos representam os valores de entropia H e complexidade C de cada imagem na base de dados. As cores correspondem aos quatro estilos (Barroco: azul, Renascimento: laranja, Pontilhismo: verde e *Hard-Edge Painting*: vermelho).

Resultados e Discussão

Após o processo de treinamento do classificador logístico, no qual cada uma das tarefas de classificação usou 80% dos dados, aplicamos o algoritmo treinado (ou ajustado) aos 20% restantes e estimamos a precisão do processo de classificação. Nesse contexto, como a cada tarefa de regressão é composta por apenas duas classes (por exemplo, Barroco ou Renascimento), a precisão do algoritmo representa a fração (ou percentual) de imagens cuja classe (estilo artístico) é corretamente identificada pelo procedimento.

Nossos resultados indicam que esse procedimento simples, baseado apenas em duas características da imagem, produz precisão razoável. No caso da distinção entre arte barroca e arte renascentista, o classificador logístico atribui corretamente o estilo da pintura em 76% das vezes. Uma precisão similar foi alcançada para classificar pinturas dos estilos Pontilhismo e *Hard-Edge Painting*, na qual a abordagem logística acerta o estilo das pinturas em 77% dos casos.

Embora distantes do valor ideal, esses resultados mostram que medidas simples e inspiradas por conceitos de Física podem ser úteis para compreender obras de arte. Além disso, essa simplicidade pode ajudar a interpretar resultados e possibilitar um diálogo com especialistas em arte. No

caso particular da entropia e complexidade de permutação, os autores da Ref. [1] argumentam que essas medidas podem ser associadas aos conceitos arte *linear* versus arte *painterly* desenvolvidos por Wölfflin [5], considerados por muitos o fundador da disciplina de história da arte. Nesse caso, a arte *linear* de Wölfflin seria representada por uma arte de baixa entropia (especialmente mais organizada), enquanto a arte *painterly* representa uma arte mais entrópica (especialmente mais difusa).

Conclusões

Neste trabalho e ao longo do projeto de iniciação científica, pudemos revisitar diversos aspectos do trabalho desenvolvido na Ref. [1] e, concomitantemente, desenvolver habilidades computacionais e teóricas associadas aos métodos de aprendizagem estatística (*machine learning*). Com relação aos resultados desse resumo, mostramos que conceitos de Física podem ser usadas para quantificar até mesmo aspectos subjetivos relacionados à arte.

Agradecimentos

Ao CNPq pelo apoio financeiro.

Referências

- [1] Sigaki, H. Y., Perc, M., & Ribeiro, H. V. (2018). **History of art paintings through the lens of entropy and complexity**. Proceedings of the National Academy of Sciences, 115(37), E8585-E8594.
- [2] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). **An introduction to statistical learning**. New York: Springer.
- [3] Ribeiro, H. V., Zunino, L., Lenzi, E. K., Santoro, P. A., & Mendes, R. S. (2012). **Complexity-entropy causality plane as a complexity measure for two-dimensional patterns**. PLoS ONE, 7(8), e40689.
- [4] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011). **Scikit-learn: Machine learning in Python**. The Journal of Machine Learning Research, 12, 2825-2830.
- [5] Wölfflin, H. (1950). **Principles of Art History: The Problem of the Development of Style in Later Art**. Courier Corporation.