Beauty Learning and Counterfactual Inference

Tao Li

Department of Computer Science

Purdue University

taoli@purdue.edu

Abstract

This work showcases a new approach for causal discovery by leveraging user experiments and recent advances in photo-realistic image editing, demonstrating a potential of identifying causal factors and understanding complex systems counterfactually. We introduce the beauty learning problem as an example, which has been discussed metaphysically for centuries and recently been proved exists, is quantifiable, and can be learned by deep models in our paper [1], where we utilize a natural image generator coupled with user studies to infer causal effects from facial semantics to beauty outcomes, the results of which also align with existing empirical studies. We expect the proposed framework for a broader application in causal inference.

1. Introduction

Recent advances in deep learning provide the capability of capturing complex data distributions and has promoted many successful applications in various domains. However, these models are mostly black boxes and deficit in explainability, demanding a timely study of "distilling knowledge from deep networks" [2]. Moreover, deep models fail to understand causal relationships behind and struggle to help answer questions such as "What if?" and "Why?" [3]. It is of great interests to not only enable automated systems to learn association rules from complicated and heterogenous data, but also be able to infer causal relations and reason counterfactually. Here, we introduce a new pipeline for causal discovery, incorporating user experiments with recent advances in image editing. In particular, we investigate the beauty learning problem, given that facial attractiveness has profound influences on multiple aspects of human society and yet is not well studied from a quantitative perspective. In the rest of the paper, Section 2 introduces necessary backgrounds in causal inference and image editing. Section 3 formalizes the problem and highlights our approach, showcasing that photo-realistic image editing along with user studies can be a powerful tool for counterfactual

reasoning. We leave more details and dicussions in the full paper [1].

2. Preliminary

Pearl's Causal Hierarchy   *Judea Pearl* [4, 5] introduced a causal hierarchy for classification of causal information, providing a theoretical foundation for analysis of causality. The causal hierarchy encompasses three levels: (i) association; (ii) intervention; and (iii) counterfactuals. Table 1 below highlights the causal ladder with examples.

Level	Typical Activities	Typical Questions	Examples
1. Association $P(y\mid x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\mid do(x), z)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\mid x', y')$	Imaging, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past two years?

Table 1. Pearl's ladder of causation [4, 5] and examples.

Structural Causal Model   Structural Causal Model (SCM) [4] lies in the key of structural causal inference. It is defined as an ordered tuple $\langle \mathbf{U}, \mathbf{V}, \mathcal{F}, \mathbf{P}(\mathbf{u}) \rangle$, where $\mathbf{U}$ is a set of exogenous variables determined by factors outside the model, $\mathbf{V}$ is a set of endogenous variables determined by factors inside the model, $\mathcal{F}$ is a set of functions that express the structures of the model, and $\mathbf{P}(\mathbf{u})$ is the distribution of $\mathbf{U}$. A diagram that captures the relationships among these variables is called a causal diagram (or graph G). With these notations, a causal effect from X to Y can be formally defined as

$$P(y\mid do(x)) \qquad (1)$$

and the effect is said to be identifiable if $P(y\mid do(x))$ can be expressed and calculated using a combination and derivatives of $\mathbf{P}(\mathbf{u})$.

Photo-Realistic Image Editing Image editing, as an important topic in computer vision, has been greatly promoted by recent advances in Generative Adversarial Network (GAN), which a system of two neural networks that contests with each other under a zero-sum game setting. It was first introduced by [6] in 2014. Since then, great progresses in both theory and practice have been made, to name a few: DCGAN [7], CycleGAN [8], and AnonymousNet [9]. Recently, *Karras et al.* [10] proposed a style-based generator architecture (StyleGAN) that is able to synthesize photo-realistic non-existent facial images which can barely be distinguished from real ones by human eyes.

3. The Example of Beauty Learning

What is beauty? It has been debated by philosophers and psychologists for centuries. The answers range from symmetry [11] and averageness [12] to personality [13] and sexual dimorphism [14]. The proverb *beauty is in the eye of the beholder* implies that perception of beauty is subjective and stems from various cultural and social settings.

3.1. Causation in Beauty Learning

Despite extensive discussions, many causal questions have never been well answered: "can a machine learn beauty semantics?" (association), "will changing a facial attribute impact attractiveness?" (intervention), and "why does this facial attribute matter?" (counterfactuals). Figure 1 provides a structural causal diagram that formalizes the question that whether attribute $\mathbf{X}$ has a causal effect on beauty score $\mathbf{Y}$ in the present of unobserved variable $\mathbf{U}$.
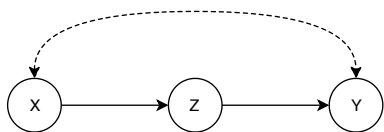

Figure 1. Causal diagram of the SCM.

3.2. Beauty Learning from Data

Previous definitions of beauty are mostly subjective and metaphysical, and deficit in accuracy, generality, and scalability. To overcome these issues, in [1] we present a data-driven study in mining beauty semantics of facial attributes, in an effort to objectively construct descriptions of beauty in a quantitative manner. We first deploy a deep Convolutional Neural Network (CNN) [15] trained by the CelebA dataset [16] to extract facial attributes. Then we investigate correlations between these facial attributes and attractiveness on two large-scale datasets with labelled beauty scores (Beauty 799 dataset [17] and US 10K dataset [18]) and accordingly select key attributes for beauty enhancements supported by statistical tests (e.g., `small nose`, `high

cheekbones`, and `femininity`). We further leverage a Generative Adversarial Network (GAN) [19] to translate these high-level representations from original images to beauty-enhanced alternatives. Figure 2 illustrates some results.
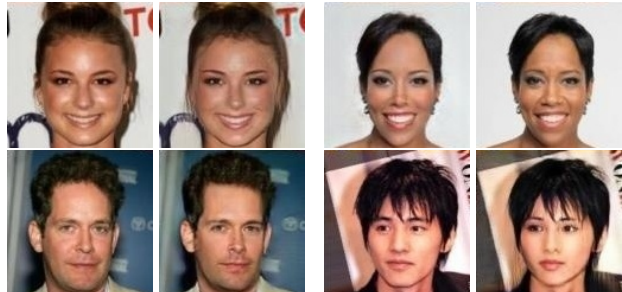

Figure 2. In each pair, left is the original image; right is the translated. Which one (left or right) do you think is more attractive?

3.3. User Study

Powered by Amazon Mechanical Turk (AMT), we perform a large-scale user study of 10,000 data points to evaluate user preferences towards difference settings, i.e., calculate $P(y|do(x))$ in the Structural Causal Model, where x is a high-level facial attribute. In this study, x has five options: `Female`, `Heavy Makeup`, `Lipstick`, `Big Nose` and `Aged`. We pick 50 celebrities from the CelebA dataset, and for each celebrity, we let the users to evaluate the original image and five translated images. Figure 3 illustrates experimental results, showing a descending order of preference level, which aligns with existing psychologist studies [20, 21].
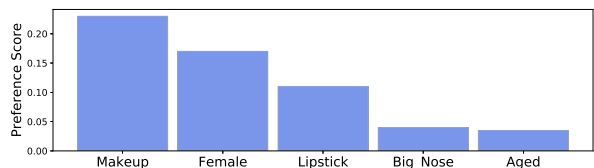

Figure 3. Results from the user experiment.

4. Conclusion

In this work, we discussed recent advances in photo-realistic image editing and their applications in causal discovery. We introduced the beauty learning problem as an example, showing that beauty exists, is learnable by machines, and can be manipulated by changing latent representations. Accordingly, we leverage a large-scale user study to counterfactually investigate causal effects from facial semantics to beauty enhancements. We anticipate more studies in the future to utilize the proposed framework as a powerful tool for causal discovery and beyond.

Acknowledgement

This extended abstract is based on [1, 22]. The author thanks colleagues at ObEN for their wonderful works.

References

[1] X. Liu, T. Li, H. Peng, I. C. Ouyang, T. Kim, and R. Wang, "Understanding beauty via deep facial features," *arXiv preprint arXiv:1902.05380*, 2019. 1, 2, 3

[2] R. Chen, H. Chen, G. Huang, J. Ren, and Q. Zhang, "Explaining neural networks semantically and quantitatively," *arXiv preprint arXiv:1812.07169*, 2018. 1

[3] J. Pearl, "The seven tools of causal inference with reflections on machine learning," Technical Report, Communications of Association for Computing Machinery, Tech. Rep., 2018. 1

[4] ——, *Causality*. Cambridge university press, 2009. 1

[5] J. Pearl *et al.*, "Causal inference in statistics: An overview," *Statistics surveys*, vol. 3, pp. 96–146, 2009. 1

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680. 2

[7] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015. 2

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232. 2

[9] T. Li and L. Lin, "AnonymousNet: Natural face de-identification with measurable privacy," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 2

[10] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *arXiv preprint arXiv:1812.04948*, 2018. 2

[11] R. Thornhill and S. W. Gangestad, "Human fluctuating asymmetry and sexual behavior," *Psychological Science*, vol. 5, no. 5, pp. 297–302, 1994. 2

[12] F. Galton, "Composite portraits," *Nature*, vol. 18, pp. 97–100, 1878. 2

[13] A. C. Little, D. M. Burt, and D. I. Perrett, "What is good is beautiful: Face preference reflects desired personality," *Personality and Individual Differences*, vol. 41, no. 6, pp. 1107–1118, 2006. 2

[14] I. S. Penton-Voak and J. Y. Chen, "High salivary testosterone is linked to masculine male facial appearance in humans," *Evolution and Human Behavior*, vol. 25, no. 4, pp. 229–241, 2004. 2

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9. 2

[16] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3730–3738. 2

[17] D. Zhang, F. Chen, and Y. Xu, "A new hypothesis on facial beauty perception," in *Computer Models for Facial Beauty Analysis*. Springer, 2016, pp. 143–163. 2

[18] W. A. Bainbridge, P. Isola, and A. Oliva, "The intrinsic memorability of face photographs." *Journal of Experimental Psychology: General*, vol. 142, no. 4, p. 1323, 2013. 2

[19] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797. 2

[20] G. Rhodes, C. Hickford, and L. Jeffery, "Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive?" *British journal of psychology*, vol. 91, no. 1, pp. 125–140, 2000. 2

[21] A. C. Little, B. C. Jones, I. S. Penton-Voak, D. M. Burt, and D. I. Perrett, "Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 269, no. 1496, pp. 1095–1100, 2002. 2

[22] X. Liu, T. Li, H. Peng, I. C. Ouyang, T. Kim, and R. Wang, "Mining semantic descriptions from data for beauty understanding," 2019. 3