# Article information:

[2011.01625] Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models  
<https://arxiv.org/abs/2011.01625>

# Article summary:

1. Shapley values are a popular method for explaining the predictions of machine learning models, but they may lead to counterintuitive explanations when features are not independent.

2. The authors propose a novel framework for computing "causal" Shapley values that take into account causal relationships between features and avoid the independence assumption.

3. The proposed method enables the separation of direct and indirect effects and can be applied to real-world examples with partial information available.

# Article rating:

Appears strongly imbalanced: The article is written in a biased or one-sided way, and the information it provides is not trustworthy enough to be considered a reliable source. You should consult other sources to find reliable information on the presented issues.

# Article analysis:

作为一篇关于解释人工智能预测的论文，本文提出了一种新的计算Shapley值的框架，称之为“因果Shapley值”。作者声称这种方法可以更好地处理特征之间的依赖关系，并且可以分离直接和间接效应。然而，在对该论文进行批判性分析时，我们发现以下问题：

1. 偏见来源：本文没有明确说明其研究背景和动机。作者没有提到任何可能存在的利益冲突或偏见来源，例如是否受到某个公司或组织的资助或影响。

2. 片面报道：本文只介绍了“因果Shapley值”的优点，但并未探讨其缺点或局限性。例如，该方法是否适用于所有类型的数据集和模型？它是否需要大量计算资源？它是否容易被误解或滥用？

3. 无根据主张：作者声称“因果Shapley值”是基于Pearl's do-calculus推导出来的，并且保留了原始Shapley值的所有优点。然而，作者并未提供足够的证据来支持这一主张。他们没有比较两种方法在不同数据集上的表现，也没有进行实验验证。

4. 缺失考虑点：本文忽略了一些重要问题。例如，在使用“因果Shapley值”时如何处理缺失数据？如何处理非线性关系？如何确定因果图中变量之间的先后顺序？

5. 主张缺失证据：尽管作者声称“因果Shapley值”可以更好地处理特征之间的依赖关系，并且可以分离直接和间接效应，但他们并未提供足够的证据来支持这些主张。他们只给出了一个真实世界例子来说明该方法如何工作。

6. 未探索反驳：本文没有探讨其他学者对该方法可能存在的质疑或反驳意见。例如，其他学者可能会认为该方法过于复杂、难以理解、难以实现、难以验证等等。

7. 宣传内容：本文似乎过于强调“因果Shapley值”的优点，并试图将其推广为一种通用解释模型预测结果的方法。然而，在实际应用中，“因果Shapley值”可能并不适用于所有情况，并且需要谨慎使用。

总之，尽管本文提出了一个有趣和有前途的想法，“因果Shapley值”，但它仍存在许多问题和局限性。我们建议作者进一步完善该方法，并进行更多实验验证和比较研究，以充分评估其可行性和有效性。同时，在撰写类似文章时，请注意平衡各方观点，并避免过度宣传自己所提出的方法或技术。

# Topics for further research:

* Potential bias sources in causal Shapley value research
* Limitations and drawbacks of causal Shapley value method
* Lack of evidence for claims about causal Shapley value method
* Unaddressed issues in using causal Shapley value method
* Insufficient evidence for claims about handling feature dependencies with causal Shapley value method
* Unexplored criticisms of causal Shapley value method

# Report location:

<https://www.fullpicture.app/item/b0e3a7f86f03633a9de4440102fba272>