A Quality Metric to Improve Scatterplots for Explainable AI

Liqun Liu[1] , Roy A. Ruddle[1], Leonid V. Bogachev[1] , Mahdi Rezaei[1] , Arjun Khara[1]

[1]University of Leeds, UK

Abstract

Scatterplots are widely utilised in Explainable Artificial Intelligence (XAI) to investigate misclassifications and patterns among instances. However, when datasets are large, overplotting diminishes the effectiveness of scatterplots. This poster introduces a new quality metric to measure the overplotting of scatterplots in the context of XAI. Initially, we assess the significance of each data point within a scatterplot by continuous density transformation, Mahalanobis Distance and a mapping function. Building on this foundation, we develop a quality metric for scatterplots. Our metric performs well accounting for rendering orders and marker sizes in scatterplots, showcasing the metric's potential to improve the effectiveness of XAI scatterplots.

CCS Concepts

• Human-centered computing → Visualisation design and evaluation methods; • Computing methodologies → Machine learning;

1. Introduction

Three primary stages of explainable artificial intelligence (XAI) are model understanding, performance evaluation & enhancement, and stakeholder communication [SSSEA20]. Within these stages, common XAI tasks aim to investigate feature importance, feature dependencies and model accuracy. Scatterplots are widely utilised in those tasks, e.g., to show SHAP values of local feature dependencies or misclassifications via t-SNE output.

A key issue with scatterplots is overplotting, especially with big datasets. Existing overplotting metrics are calculated from either the values of data that is plotted (e.g., [SNLH09]) or images of the plot that is created (e.g., [TAE*09]). However, few metrics take account of how important different visual patterns are for the insights users gain in XAI or visual factors such as marker size.

Our approach addresses the gap by introducing a quality metric for scatterplot overplotting that considers the significance of each data point, data point coordinates and visual factors. This poster describes the metric and evaluates them with two XAI classification model scenarios.

2. Related Works

We categorise quality metrics according to the number of variables and measurement approach. The number of variables is either two (X and Y) or three variables (X, Y and a category encoded using colour or shape). Regarding the measurement, metrics either are based on rendered images or coordinate data from the scatterplot. This bifurcation yields four groups: 1) XY scatterplots based on coordinate data, such as Pearson's coefficient [MPOW17]; 2) three-variable scatterplots measured with coordinate data, such as the Distance Consistency [SNLH09]; 3) XY scatterplots through rendered images, like the Rotating Variance Measure [TAE*09]; and 4) three-variable scatterplots assessed via rendered images, including Class Density Measure [AEL*10, TBB*10].

3. Method

We propose a quality metric that initially evaluates the significance of each data point's impact on observers' insights. Subsequently, we quantify the extent of information that remains obscured for three-variable scatterplots.

3.1. Data Point Significance

We employ an approach to determine the significance of data points with three steps: 1) transforming the data points into a continuous density representation in arrays, effectively capturing the data information at the pixel level. The transformation relies on the markers' size and shape, by quantifying the extent of pixel coverage by each marker; 2) calculating the significance of each data point by Mahalanobis Distance, which assigns a distance to each data point based on its position relative to a specified distribution, involving either a cluster or a correlation pattern. A greater distance indicates a higher significance of a data point, suggesting that it is further from the central tendency of the distribution, and 3) mapping the significance to fall within the range of 0 and 1. We employ a two-step method to normalise these values, which includes initial normalization followed by a power function to enhance distinctions.
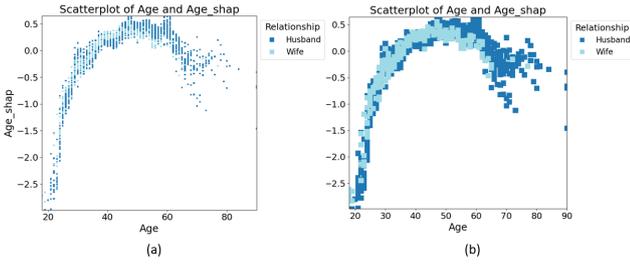
Figure 1: *The sensitivity of quality metric to marker sizes. Scatterplots rendered with varying marker sizes illustrate the impact of marker size on our quality metric. (a) achieves a quality metric score of 0.76. In contrast, scatterplot (b) scores lower at 0.41.*

Figure 2: *Sensitivity of the quality metric to rendering orders. In (a), the digits 2, 5, and 8 are rendered before the others, while in (b) the three digits are rendered after all others, leading to different quality metric scores of 0.41 and 0.61 respectively.*

3.2. Quality Metric

In this section, we introduce the principles of our proposed quality metric, which represents how much information remains obscured within a scatterplot. The foundational concept of the metric involves quantifying the activation of pixels—specifically, the count of activated pixels and the frequency of their activation—integrated with the weight values (significance values). The final result is a number between 0 and 1, where a higher value indicates less information is hidden.

In a scatterplot, consider a pixel as $k \in K$, where K is the set of all pixels. Each pixel k is overlaid by N data points, and each data point, denoted as m_i where $i \in \{1, 2, \ldots, N\}$, within this pixel each data point carries a weight value shown as w_i^k, calculated using Mahalanobis Distance and mapping function. Data points within each pixel are rendered in sequence from $i = 0$ to $i = N$. The data point at $i = N$ is referred to as the top-layer data point. We calculate the top-layer information by the weight value of a top-layer data point for a pixel k, shown as $q_t^k = w_{i=1}^k$.

Beneath the top-layer data points lie hidden data points, which correspond to hidden information. The hidden information for a pixel k is composed of two parts, depending on whether the hidden data points belong to the same class as the top-layer data point. These parts are calculated using Eq. (1) and (2):

$$q_s^k = \frac{\sum_{i=1}^{N-1} w_i^k \cdot \mathbb{1}\{c_i^k = c_t^k\}}{\sum_{i=1}^{N-1} \mathbb{1}\{c_i^k = c_t^k\}} \quad (1)$$

$$q_d^k = \sum_{i=1}^{N-1} w_i^k \cdot \mathbb{1}\{c_i^k \neq c_t^k\} \quad (2)$$

Here, q_s^k represents the hidden information when hidden data points are in the same class as the top-layer data point, while q_d^k accounts for those in different classes. The $\mathbb{1}\{\cdot\}$ notation denotes the indicator function, yielding 1 when the condition is true, and 0 otherwise. The variable c_i^k denotes the class of the i-th data point on pixel k, and c_t^k denotes the class of the top-layer data point. Our overall quality metric, combining these elements among all the pixels K in the scatterplot, is formalised in Eq. (3):
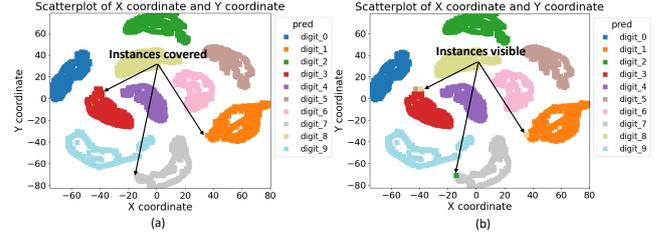
$$Q = 1 - \frac{\sum_{k \in K} q_s^k + \sum_{k \in K} q_d^k}{\sum_{kinK} q_s^k + \sum_{k \in K} q_d^k + \sum_{k \in K} q_t^k} \quad (3)$$

4. Results

We assessed the effectiveness of the quality metric proposed by generating scatterplots with varying marker sizes and rendering orders to determine the sensitivity of the metric to visual factors.

The first example employed an XGBoost model to classify adult income dataset [AN07], with the scatterplots to visualise SHAP values, which are utilised to examine the sensitivity of our quality metric to variations in marker sizes. Fig. 1 showcases two scatterplots with marker sizes set to 5 and 120. As the marker size increases, a greater number of data points labelled *Husband* are obscured by those labelled *Wife*, leading to a corresponding decrease in our quality metric scores from 0.76 to 0.41. This outcome aligns with the expectation that increasing marker size can aggravate overplotting in scatterplots.

The second example utilised a CNN model for classifying the MNIST dataset [Den12], with results visualised in scatterplots where each point represents a single instance (see Fig. 2). The two scatterplots render the data in different orders and demonstrate how that influences our quality metric scores. Certain critical instances that are obscured in (a) are pivotal for AI experts to identify model errors. Our quality metric assigns scores of 0.41 for (a) and 0.61 for (b), indicating that scatterplot (b) hides less information.

5. Limitations and Future Work

Despite the strengths of our quality metric, it has limitations, such as its lack of consideration for the opacity of scatterplots. Although both examples originate from the XAI field, we believe our quality metric is versatile and can be applied to scatterplots in various other domains.

Acknowledgements

References

[AEL*10] ALBUQUERQUE G., EISEMANN M., LEHMANN D. J., THEISEL H., MAGNOR M.: Improving the visual analysis of high-dimensional datasets using quality measures. In *2010 IEEE Symposium on Visual Analytics Science and Technology* (Salt Lake City, UT, USA, Oct. 2010), IEEE, pp. 19–26. URL: http://ieeexplore.ieee.org/document/5652433/, doi:10.1109/VAST.2010.5652433. 1

[AN07] ASUNCION A., NEWMAN D.: UCI machine learning repository, 2007. 2

[Den12] DENG L.: The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine 29*, 6 (Nov. 2012), 141–142. Conference Name: IEEE Signal Processing Magazine. URL: https://ieeexplore.ieee.org/abstract/document/6296535, doi:10.1109/MSP.2012.2211477. 2

[MPOW17] MICALLEF L., PALMAS G., OULASVIRTA A., WEINKAUF T.: Towards Perceptual Optimization of the Visual Design of Scatterplots. *IEEE Transactions on Visualization and Computer Graphics 23*, 6 (June 2017), 1588–1599. URL: http://ieeexplore.ieee.org/document/7864468/, doi:10.1109/TVCG.2017.2674978. 1

[SNLH09] SIPS M., NEUBERT B., LEWIS J. P., HANRAHAN P.: Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum 28*, 3 (2009), 831–838. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2009.01467.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2009.01467.x, doi:10.1111/j.1467-8659.2009.01467.x. 1

[SSSEA20] SPINNER T., SCHLEGEL U., SCHÄFER H., EL-ASSADY M.: ExplAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2020), 1064–1074. doi:10.1109/TVCG.2019.2934629. 1

[TAE*09] TATU A., ALBUQUERQUE G., EISEMANN M., SCHNEIDEWIND J., THEISEL H., MAGNORK M., KEIM D.: Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. *VAST 09 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings* (2009), 59–66. ISBN: 9781424452835. doi:10.1109/VAST.2009.5332628. 1

[TBB*10] TATU A., BAK P., BERTINI E., KEIM D., SCHNEIDEWIND J.: Visual quality metrics and human perception: an initial study on 2D projections of large multidimensional data. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Roma Italy, May 2010), ACM, pp. 49–56. URL: https://dl.acm.org/doi/10.1145/1842993.1843002, doi:10.1145/1842993.1843002. 1