The Comparison of Glyph Based Visualization Techniques For Visualizing Feature Vector Similarity

James Carr-Pries
20769280
Systems Design Engineering
Department of Systems Design Engineering, University of Waterloo

Abstract – The goal of this report is to determine the best feature vector visualization method for a visual understanding of the similarity between instances within a tabular dataset. Image-based datasets are very powerful and the property of the data being images is used by researchers to create impactful visualization tools. Tabular data is unable to utilize these visualization tools because there is not an effective way to view a feature vector of several numbers. This research compares two existing visualization techniques – Star Glyphs and Chernoff Faces – against a proposed solution – Reticle Glyphs – to determine the best technique for visualizing similarity. Visualizing similarity is important to help enhance the visualization tool I created this past term. To test and compare these methods a survey was conducted, and interviews were held with Machine Learning Scientists within Layer6. The results of the study show that the proposed solution serves as the most accurate visual representation of similarity within the limitation of the research. Further research must be conducted to determine why the Reticle Glyphs does not garner the highest confidence from its users.

Keywords – Reticle Glyphs, Data Visualization, Explainable AI

I. SITUATION OF CONCERN & PROJECT OBJECTIVES

The overarching goal for my fourth co-op term, at layer 6, was to develop a visual tool within the company's existing web application that would empower machine learning scientists (MLS) to be able to explore the embedding of their data. This task was to be carried out independently with guidance and feedback from my mentor and other MLS. In order to achieve this goal, I developed an interactive scatter plot of the data. This tool can be broken down into five main components: filtering, embedding extraction, dimensionality reduction, visualization, and clustering.

First, the data can be filtered based on one of many criteria that the MLS needs. These filters included, the data set (training or test), the data range, feature value range, prediction percentile or even a combination of these criteria. This data is then retrieved from an Elasticsearch server.

The embedding of the model is a compressed or encoded version of the input data. In neural networks, a hidden middle layer could be used as an embedding. Although, at Layer6 all the models use a method called XGBoost. An MLS at Layer6 developed a technique that leverages the physical tree structure of the model and truncated simple vector decomposition to create a data embedding for XGBoost models. I first performed this embedding extraction technique to find a data embedding for the visualization tool.

This embedding technique was able to reduce the dimensions of the input data from over 250 dimensions to 20-50 dimensions. However, to be interpretable to the MLS using the tool the embedding had to be reduced further to two dimensions. To achieve this reduction, I performed the MLS's choice of t-SNE or UMAP. These two techniques are well established in the industry and use manifold learning to create a non-deterministic mapping of the data from higher dimensions into 2- or 3-D.

After this dimensionality reduction is performed, the data is then sent to the front end where it is displayed using D3. In order to enhance the interactivity, I developed a way to dynamically display the points stored in a KD-tree on the screen as the user pans and zooms. There was also additional information about a cluster or a point available in a preview area to the right of the plot when a point was clicked on by the user.

The dimensionality reduction technique naturally reveals clusters based on the geographical layout of the points on the screen but in order to provide an additional layer of information. I chose to overlay colour-based cluster information on the data points. The raw data was clustered using the hierarchical clustering technique using the feature values or the SHAP values. The SHAP values are values given to each feature of every instance between 0 and 1 that denotes that feature's contribution or importance toward that instance’s prediction. Clustering based on SHAP values is a common explainability technique and leads to improved clustering because all the features are mapped to the same unitless feature space. After the clustering, each cluster is assigned to a colour and then the points are coloured accordingly. This added layer of information can help MLS check that the clusters that come from the model embedding align with the clusters formed through colour as a kind of second check. Additionally, the MLS can overlay clustering based on a different feature to provide an additional layer of information.

Once the first version of the tool was completed, I got feedback from the users. This feedback occurred through eight cognitive walkthroughs in early April. From these interviews, I uncovered that there was a need to have a further understanding of the underlying data and how points related to one another. Clicking each point and parsing through the information available in the preview area was unintuitive and non-immersive as they had hoped; they needed a way to view that information within the visualization.

This was the initial motivation for my glyph-based similarity technique. However, there was another motivation with an equal weight that led this research project. Throughout the term, I was a part of a visualization reading group that investigated various machine learning visual interface design research papers [1]. Many of the useful techniques leveraged the fact that the data being analyzed was image-based or video-based data [1]. This classification of data provides the user of the visualization tool an extra level of understanding and interpretability because they can assess the properties of the data instance and use innate visual perception facilities to compare it to other instances. This extra visual understanding is not readily available when working with tabular data that can consist of feature vectors with hundreds of features.

The personal motivation from the tool I developed throughout my co-op term and the motivation stemming from the visualization research lead to my research question: what feature vector visualization method allows for the best visual understanding of similarity for tabular data? Similarity can be measured mathematically, however, many mathematical metrics can be used to measure similarity such as Euclidean distance, Shannon entropy/manifold learning techniques, Minkowski distance and cosine similarity. Cosine similarity was chosen to measure similarity in this research because it was within my mathematical understanding but slightly more accurate when working in high dimensions than Euclidean distance [1]. The cosine similarity is a measure of similarity between two feature vectors of an inner product space:

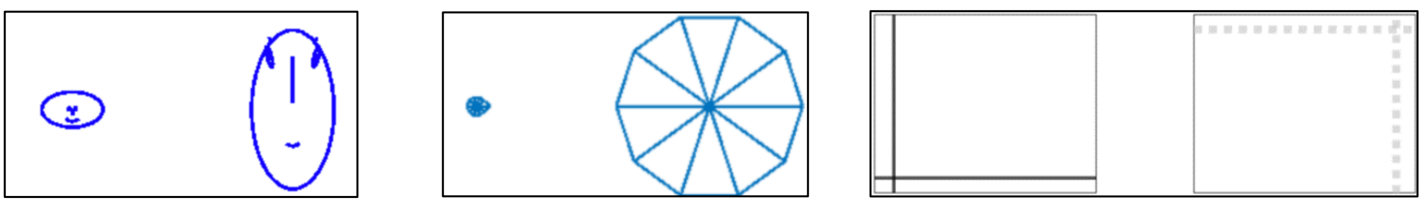
$$\text{similarity} = \cos\theta = \frac{\boldsymbol{A} \cdot \boldsymbol{B}}{||\boldsymbol{A}||\,||\boldsymbol{B}||} \tag{1}$$

The effectiveness of the design techniques at visualizing similarity is measured by comparing existing techniques to a proposed solution using master-apprentice interviews with industry experts and a survey based on notable visualization evaluation methods. The success of the technique is determined by a higher score in the survey, higher confidence in the technique and support from the interviewees. The survey score will be an assessment of how well the visualization technique visualizes cosine similarity.

II. DESIGN METHOD AND ENGINEERING ANALYSIS

Design Method

There are two main parts to my research: designing an alternative method for visualizing similarity and creating a survey to test the methods and perform engineering analysis to evaluate the performance concerning the research objective. The first step for designing the alternative design technique was to do a prior works search to find pre-existing work in the field. Through this work, I found two main methods and several variations on those methods [2]-[6]. Each of these methods is glyph-based. In this context, a glyph is a small visual object that can be used to depict attributes or the composition of a data set [2]. Similar but unique from an icon or a symbol. The first, and more prominent method is Star Glyphs (Figure 1b) and the second is Chernoff faces (Figure 1a). In both methods feature values are first mapped to a linear scale and then mapped to the corresponding property of a glyph. For the Star Glyphs, illustrated in Figure 1a, 10 features are matched to a specific axis. The longer the axis the larger the feature value it represents up to that feature's maximum in the data set. For the Chernoff faces, illustrated in Figure 1b, instead of being mapped to an axis, the feature values are mapped to properties of the face such as the size, distance between the eyes or the length of the nose. Both methods are used to convey the meaning of the underlying data in a glyph form [5][6]. For example, when looking at a star glyph it is possible to approximate the numerical values it represents. However, neither of the designs was developed to visualize the similarity between points [2]. Additionally, the other designs are more powerful when analyzing raw data but do not leverage the benefits of visualizing a model embedding [5][6].



(a) Chernoff Faces (b) Star Glyphs (c) Reticle Glyphs

Figure 1: Comparison of three different glyph-based data visualization techniques with the minimum (left) and maximum (right) value for each of the 10 features [Image Source: JCP, 2021]

The design I created, Reticle Glyphs, illustrated in Figure 1c, is specific for the use in machine learning when visualizing an embedding. The Reticle Glyphs are unique because they map the top 10 most important features to the most important visual variables. Visual variables are the foundational properties of a visual mark that can be adjusted in order to create a more meaningful image [7]. Examples of visual variables are length, position, thickness, orientation, colour, or shape. Using existing visualization research, I created a ranking on which visual variables convey more prominence in human perception: Position > Value (darkness) > Thickness > Grain (Pattern) [8][7]. The only value, which is the darkness of the shade of grey, is used instead of colour because colour cannot be measure or ordered [8]. Colour tends to be overused in visualization but in practice, there is no evidence to suggest users can measure the difference between red and yellow or order the colours blue, red and green [7]. In addition, colour is less accessible to users since 8% of men are affected by colour blindness [8].

Once I completed the research of visual prominence, I was easily able to match the most important feature to the most prominent visual variables. The most important features are determined using Shapley values. Shapley values are an Explainable AI technique that leverages concepts from game theory to determine the "payout" (importance) for each "player's" (feature's) contribution to the "outcome" (prediction). The hypothesis was that by carefully mapping the visual variables of the glyph to corresponding features that there would be a more natural interpretation of similarity when analyzing points within a model embedding. Both the glyph and the embedding would convey the meaning of the underlying data and the structure of the model [4].

Engineering Analysis

The second part of my project was to determine a method to analyze and compare the techniques to determine which method was best according to the specified criteria: visualizing cosine similarity. Many perception-based surveys exist within the field that has been used to determine the effectiveness of designs and how well icons or glyphs are at conveying their intended information [2][4]-[6][8]. Two components are present in most of these surveys: response time and question confidence. For surveys that are looking to test quick responses or pop-out effects, it is important to include a variable of time [8]. However, this variable was omitted due to the limitations of online surveys. These limitations will be addressed further in the limitations section. The confidence measurement was incorporated into my survey. The confidence metric is a question following each evaluation task that asks the respondent how confident they are in their answer. This is important because when using these glyphs in visualization tools such as the embedding visualization tool I created there will not be a score at the end that articulates how well and frequently the glyph was used. The effectiveness and reliability of the visualization are correlated to how much trust the user has in the visualization to convey the information it is supposed to convey [6]. If the visualization is not trusted it will not be factored into the visual perception when using the tool [6].

An author that compared different visualization techniques addressed the importance of both global and local questions. Global questions require the context of the entire dataset. Global questions can help solve the task of ordering points in terms of similarity [6]. However, local questions only require the consideration of a few points and can be useful in determining the effectiveness of spotting differences within a cluster or an outlier that does not belong to the cluster [6]. To be an effective visual aid in the embedding visualization tool I created the glyph technique must be effective in both determining class membership in the global context and the variation within a class (local).

The survey I created tests how effective three unique glyph techniques are at visualizing similarity. The three glyph techniques tested in the survey are Chernoff's Face, Star Glyphs, Reticle (crosshair) Glyphs. For each technique, there were 12 questions. For each question, the respondent was asked to rate their confidence in their answer from 0-5. Four questions have four glyphs displayed and the respondent must choose the one that is least like the others. Four questions have a prototype glyph and asked the respondent to choose one of the glyphs displayed that is most similar to the prototype. Four of the questions asked the respondent to order the three glyphs displayed in terms of similarity to the prototype. These answers to these questions, the type of questions and the glyph type were all randomized to limit confounding effects.

The final form of evaluation was through the facilitation of master-apprentice online interviews. This was an open interview where I described each of the techniques to the MLS and allowed them to share their thoughts and reactions to each of the designs. I also surveyed two researchers that specialize in visual interface design. I was able to gain feedback based on their heuristics. The three guiding questions that I used as prompts throughout the interview were:

1. How well does each of the techniques convey a visual representation of similarity?
2. What could be changed to help achieve a better visual representation of similarity?
3. What existing features of this design help convey a visual representation of similarity?

The survey was distributed using Google Forms and the data analysis was all performed within Google Sheets. The effectiveness of each design technique was determined by assessing the question score and the question confidence. For the first question which asked the respondent to choose the glyph that did not belong there was one point awarded for the correct answer and 0 for the wrong answer. For the second type of question where the respondent had to determine the most similar glyph to the prototype, there was one point for the correct answer or a score from 0-1 based on how similar the selected answer is to the correct answer. The least similar option is 0 points. For the third type of question, there were 0.25 points for every correct position and 0.125 points for any position 1 off from where it should have been. For example, if the correct matching was A-Closest, B-2nd, C-3rd and D-4th but the respondent had A-Closest, D-2nd, B-3rd and C-4th they would receive 0.25 points for a correct answer for the closest, 0 points for 2nd closest because it was supposed to be 4th (over 1 position away) and 0.125 points for 3rd and 0.125 points for 4th. After all, B and C's correct answers were only one position away. This scoring system was used to ensure that answers that were close to being right were not completely deemed inaccurate because when these visualizations are being used in a visualization tool the precision is more important than just being correct.

III. RESULTS

The survey results are broken down into three sections for each of the glyph techniques. For each section, there are 12 questions broken into three categories: three local questions, three global ordering questions and three global pick-the-closest questions. Additionally, for each question, there is a confidence score. Table 1 is a record of the mean and standard deviation for the question score (between 0-1) and the question confidence (between 0-5). Figure 2a illustrates the average category score for each technique where the category score was the sum of all the points within the technique (between 0-4). Figure 2b illustrates the average category confidence for each technique where the category confidence was the average of each of the confidence ratings within the category. The feedback received from the interviews will be incorporated into the discussion section to support or oppose the data presented in this section.

Table 1: Survey Statistics for the question score and question confidence for each of the three glyph-based visualization technique [Data Source: JCP, 2021]

	Mean Score	Score Std.	Mean Confidence	Confidence Std.
Star Glyphs	0.5967	0.230	3.120	0.594
Chernoff Faces	0.4958	0.187	1.837	0.486
Reticle Glyphs	0.7375	0.247	2.763	0.701

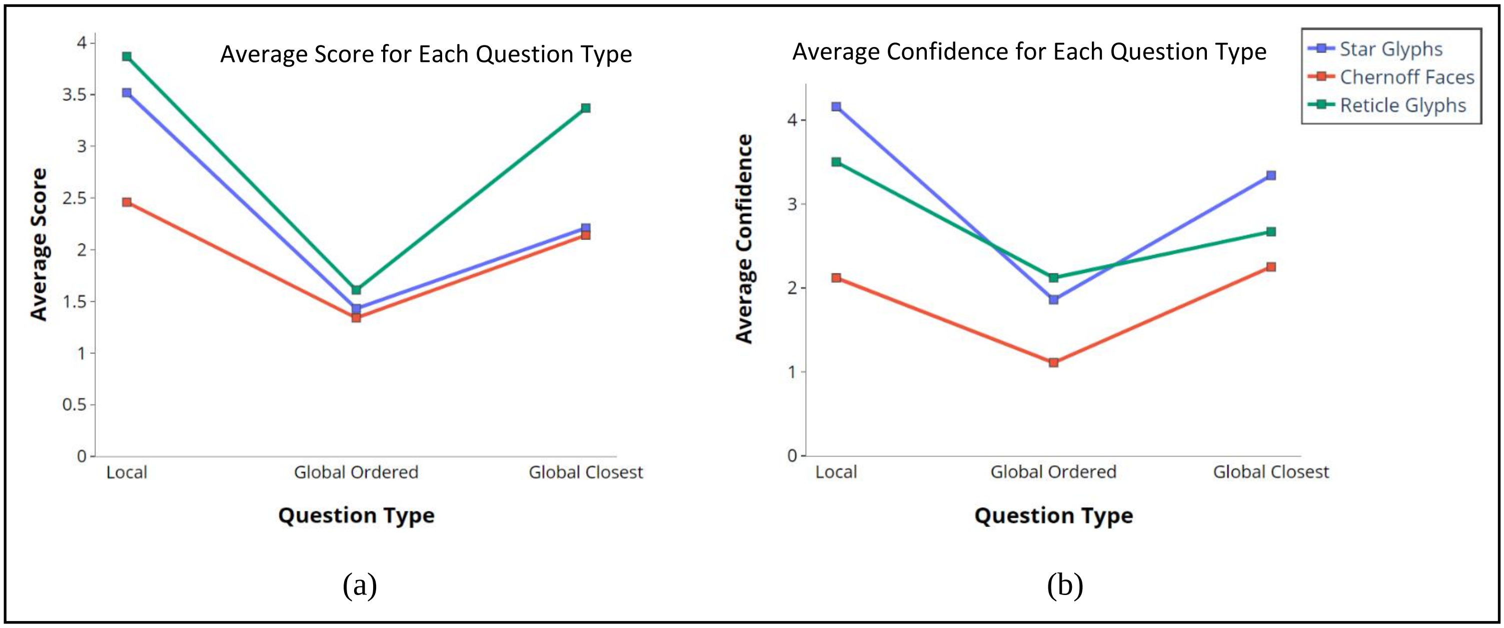


Figure 2: Average Score and Confidence for each question type [Image Source: JCP, 2021]

IV. DISCUSSION AND DESIGN VALIDATION

Several observations can be made from the data collected. The first and most important conclusion to the original question asked is that based on Table 1 the Reticle Glyphs have the highest mean score and based on Figure 2a it has the highest average score in each question category. This is important because the original question was what feature vector visualization method allows for the best visual understanding of similarity for tabular data based on cosine similarity? Although there are other factors to consider it is evident from the results that the Reticle method conveys the most accurate visual representation of similarity.

Additionally, The Chernoff faces resulted in the lowest score and confidence for each of the categories and overall. For the interviews, it was a general acknowledgement that this technique was very difficult to interpret. One interview said, “it is difficult to decipher which facial feature corresponds to which data feature and it is even harder to determine small changes within that facial feature.” Other interviews commented that the faces were too distracting and specific features of the face drew a disproportionate amount of attention. The data and the interview responses showed that Chernoff Faces is the worst out of the three glyph methods for visualizing cosine similarity.

Another observation that should be noted is the confidence score. For each of the methods, the confidence dropped significantly for the questions that asked the users to order the results. This shows that respondents were not confident when having to make precise or detailed comparisons between the glyphs. This is expected because these glyphs are not meant to be used for critical analysis but rather to support the interactivity within a visualization tool like the embedding visualization tool I developed. Within this tool, there are more fine-grained details available that would be able to provide support for challenging comparisons of several points.

The confidence score for Reticle Glyphs was surprisingly lower than that of the Star Glyphs despite the Reticle Glyphs having a higher score. Several factors could be at play here and this may require further research. However, comments from the interviews and prior research shed a light on the matter. There was a general sentiment that the simplicity and straightforwardness of the Star Glyphs were familiar. This simplicity may have caused an unwarranted sense of confidence in the results because the respondent thought they had a better understanding of the glyph's implementation details. The lower confidence score of the Reticle Glyphs may also be a result of its implementation details. These glyphs were created based on research that leverages the subconscious visual perception abilities of the respondents [8][7]. The user may not be aware that the feature that they think is more important is meant to be more important. This is opposed to the Star Glyphs which requires visual recognition and analysis to interpret the glyph more cognitively. The sense of being unaware of the implementation-specific of the Reticle Glyphs could result in lower confidence in the method.

V. LIMITATIONS OF METHODS USED AND DESIGNED SOLUTION

Sample Size

There were three main limitations to the design and the design analysis: the sample size, the similarity metric, and the lack of in-person surveys. The sample size for both the survey and the interviews was limited to members of the ML Systems team or similar teams within the organization. This is a limitation because a smaller sample size results in less confidence when generalizing sample statistics to population parameters. This is because a smaller sample is more susceptible to effects of randomness and can result in inconclusive results. A smaller sample size would also help explain the relatively large standard deviations for each of the measurements calculated in Table 1. As the sample size decreases the variance within the data increases. Although a larger sample would have resulted in a higher degree of confidence, I do not think it would impact the discussion because the general results showed a strong discrepancy between the different visualization methods.

Another factor to consider with a small sample size is an underrepresentation of certain population characteristics. For example, no one who participated in my research had visual impairments like colour blindness and 95% of the participants were male. However, this limitation should not affect the results of this research because there is no evidence to support visual perception differences in these recognition tasks between males and females [8]. Additionally, all the research that formed the basis of the Reticle Glyph implementation was impairment agnostic which means that those with colour-blindness should not have different results [8][7]. For example, the use of colour as a visual variable in the glyphs was omitted and the lightest shade remained was limited to within a distinguishable range while being presented on a white background [8].

Similarity Metric

The second limitation of the design is the use of cosine similarity as the similarity metric. This metric was chosen because it was within my mathematical understanding. As a professional engineer in training, I must practice the engineering code of ethics which states that I should not use methods that I do not understand, and I should not falsely proclaim knowledge on a subject that I do not understand.

The reason using cosine similarity is a limitation to the design is because this similarity metric does not perform as well on high dimensional data across multiple units of measurement [9]. For example, it is not logical to compare units of dollars to another feature in units of length. In order to address this impactful limitation, I performed a whitening operation that essentially converts all the measurements to a standard normal distribution, so all the units are in standard deviations. Although this whitening operation fixes the main limitation, other limitations are still present after this operation. When performing this operation there is the assumption made that each feature takes on a normal distribution which is not necessarily the case for all features.

Finally, on a broader scale, using the cosine similarity metric to measure similarity between high dimensional points has proven to have limitations because when working in higher dimensions the data points tend to fall onto certain manifolds or hyperplanes and are not evenly spaced out [9][10]. The easiest way to illustrate this is to use a piece of paper within a three-dimensional contained volume folded into a U shape. At higher dimensions (~100-D) research has shown that points will fall on this hyperplane that is being represented by the folded paper [10]. Points that are similar to one another will lie beside each other on the piece of paper. The cosine similarity method would state that two points, one on each of the tips of the U, would be similar to one another because the Euclidean distance between those points is relatively small. Yet, manifold learning research would suggest that they are as dissimilar as they could be in this example because the distance when travelling along the manifold – the face of the sheet of paper – the distance is relatively large [10]. In my research, I am only working on data in 10 dimensions so manifold-based similarity measurement techniques would provide more accuracy but not dramatically affect the results or the conclusions of those results.

Online Surveys

The final limitation of the design analysis technique is the use of online surveys. Online surveys are harder to regulate and are less flexible. Past visualization research surveys that compare different methods ensure that the viewing conditions for all of their surveys are consistent so that external environmental variables do not contribute to the results of the survey [5][6]. When working online it is less feasible to ensure that all participants have the same screen brightness, the distance between their face and the screen, lighting conditions, view background, etc.… All of these environmental factors can affect the results of the survey and so it is important to control these as much as possible [5]. For example, having a brighter screen can help the user more easily discriminate different shades and amplify certain features greater than others whereas a darker screen could mask small differences in shades. To limit most of the impact of one of the more impactful variables, distance from the screen, all of the users were asked to maintain one arm’s length away from the screen. Although this guideline does not fix all of the issues with the viewing conditions it provides more consistency across the responses.

The other limitation involved in using an online Google Forms survey is the inability to time how long the user takes at answering each question. This does not affect the current results or the discussion, but it prevents a deeper assessment of the design and its ease of use. Having this data would have helped determine which method resulted in a quicker interpretation and understanding which would make the ease of use of the final product better and more natural.

VI. CONCLUSIONS

The objective of this report was to determine what feature vector visualization method allows for the best visual understanding of similarity for tabular data? In order to achieve this objective, the Reticle Glyph visualization was purposed to challenge existing glyph-based visualization methods. The survey and the interviews clearly concluded that the Reticle Glyph method was the best at viewing cosine similarity between feature vectors. This result supports the research upon which the Reticle Glyph method was founded and shows a successful implementation. This method however did not perform as well in terms of confidence because it used less explicit methods of conveying feature values and relied heavily on implicit perception-based research. The Reticle Glyph was designed to leverage properties of visual perception instead of visual recognition and interpretation to create a more natural and seamless experience when embedded within the visualization tool. Further research should be conducted to determine the speed of use of the Reticle Glyphs method.

VII. RECOMMENDATIONS

The primary recommendation to uncover a deeper understanding of the results and better address the situation of concern is to develop a survey technique that is capable of timing the response duration for each question for the Star and Reticle Glyphs.

One of the main limitations of the design analysis method used in this report was the lack of flexibility available from using an online Google Forms survey. This limited the capabilities of the survey and prevented the ability to time the duration of each question. When considering the effectiveness of the visualization technique it is important to be able to see how long users take from when they are shown the glyphs to when they can make an appropriate action based on that image [6]. This is especially important when embedding this glyph into a visualization tool that is already conveying information to the user [6]. If the visualization method is to be effective and beneficial within the context of the tool it should be easy and quick to interpret and decipher the correct course of action [6].

There are several costs associated with this recommendation. The first is the time required to find or create a survey tool that allows for timed questions. This should not be too costly because tools like D2L Learn Quizzes already have features like that and there are many extensions to Google Forms that could be added to enable this capability. Another cost is the cost of re-distributing the survey. The first time the survey was distributed it took users less than 20 minutes to complete and all the results were collected within one day. With a timer feature added the test times will likely be quicker. This time cost is distributed between all participants, so it is not costly. The final cost is to recompile and analyze the results. The structure for determining the score and computing the results is already in place so the only cost will be the additional work of collecting the times for each question and computing those results. This should only take one or two days of work. In total, the cost of this recommendation is approximately one week of work for a full-time employee.

One week of work is a small cost because implementing this recommendation will help solidify the results previously discussed and determine which method is more useful when working in the context of a visualization tool. It is important to determine the speed of response for each technique and compare it to the confidence metric for each technique. From the result of the survey, it was shown that the Reticle Glyphs resulted in a more accurate but less confident response. Determining the speed at which users can answer the questions will help determine whether the lack of confidence was attributed to the use of implicit visual perception techniques that are more useful with quick recognition rather than in-depth analysis.

An optional second recommendation to help confirm the results of the survey is to add the glyphs to the embedding visualization tool and conduct interviews with MLS. The interview would take on the form of an in-person cognitive walkthrough so the MLS would perform certain tasks or work to achieve certain objectives. These in-person interviews would require a greater cost because they will take longer to perform and will require more critical thinking and analysis. Additionally, there would be a development cost associated with implementing the glyph method into the tool. An estimated time for this work is 2-3 weeks. This interview style analysis is important, and the results would outweigh the costs because these interviews could give more concrete real-world results that could be compared to the results of the revised survey.

ACKNOWLEDGMENTS

I would like to acknowledge the help and support that I received from Barum Rho, KK Leung, and Kyle Hall. Barum was my supervisor and supported me along the development process and provided feedback and guidance throughout. KK Leung helped me understand many of the mathematical techniques and implementation work. Kyle Hall helped guide the visualization research and provided guidance with the creation of the survey. Additional acknowledgement is given to the 20 survey participants from Layer6 and the other members of the ML Systems team for their help throughout the term.

REFERENCES

[1] C. Chen et al., "OoDAnalyzer: Interactive Analysis of Out-of-Distribution Samples," *IEEE Transactions on Visualization and Computer Graphics*, 2020. doi: 10.1109/TVCG.2020.2973258.

[2] R. Borgo, J. Kehrer, and D. H. S. Chung, “Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications,” *Eurographics State of the Art Reports*, 2013.

[3] E. Fanea, S. Carpendale, and T. Isenberg, “An Interactive 3D Integration of Parallel Coordinates and Star Glyphs,” *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS'05)*, May 2005.

[4] G. Dzemyda, O. Kurasova, and J. Žilinskas, *Multidimensional Data Visualization. Methods and Applications.* Springer, 2012.

[5] M. D. Lee, R. E. Reilly, and M. A. Butavicius, *An Empirical Evaluation of Chernoff Faces, Star Glyphs, and Spatial Visualization for Binary Data*. Jan. 2003.

[6] I. I. Kamsani, N. Ariff, and N. Khalid, “Survey of Star Glyph-Based Visualization Technique for Multivariate Data,” *Australian Journal of Basic and Applied Science*, vol. 9, no. 26, pp. 61–66, 2015.

[7] M. S. T. Carpendale, “Considering Visual Variables as a Basis for Information Visualization,” *Department of Computer Science*, Jan. 2003.

[8] P. T. Quinlan and G. W. Humphreys, “Visual search for targets defined by combinations of

color, shape, and size: An examination of the task constraints on feature and conjunction searches," *Perception & Psychophysics*, vol. 41, no. 5, pp. 455–472, 1987.

[9] K. Olobatuyi, "Cluster Weighted Model Based on TSNE Algorithm for High-Dimensional Data," 2021.

[10] "Large-Scale Manifold Learning," Manifold Learning Theory and Applications, pp. 145–168, 2011.

Sources of Images - Images presented in this document created by the authors are indicated by the authors initials and the year of creation