Unsupervised clustering applied to the optimization of a Case-based Reasoning system for the selection of optimal image explanation methods

Esteban E. Brito-Borges[1,2], Mauricio G. Orozco-del-Castillo[1,2,*] and Juan A. Recio-García[3]

[1]*Tecnológico Nacional de México/IT de Mérida, Department of Systems and Computing, Merida, Mexico*

[2]*AAAIMX Student Chapter at Yucatan, Mexico (AAAIMX), Association for the Advancement of Artificial Intelligence, Mexico*

[3]*Department of Software Engineering and Artificial Intelligence, Instituto de Tecnologías del Conocimiento, Universidad Complutense de Madrid, Spain*

Abstract

The goal of this paper is to develop a cluster-based retrieval process to select the optimal explanation method for a given image and its corresponding classification by a neural network model. We propose the use of a density clustering method to organize a case base consisting of images labeled according to their optimal explanation method. This approach presents a prediction accuracy similar to a standard nearest-neighbor method, but significantly reducing the required retrieval time.

1. Introduction

It is widely known that artificial intelligence (AI) systems are being used in many areas of industry today. One area of AI in particular, machine learning (ML), is preferred because of the performance of its models, based on statistical learning. However, many of this models are considered as "black boxes", because their internal processes are difficult to interpret with respect to the predictions and outputs they produce [1]. Solving this problem is a requirement to audit the reasoning behind incorrect decisions taken by AI systems, preventing or reducing the problems these decision could carry, like the ones presented in companies such as Google and Facebook related to offensive image misclassifications [2].

Nowadays, eXplainable AI (XAI) techniques help on the internal understanding of black box models by following two main approaches: (i) model-dependent techniques that can only be applied to a specific type of ML model (because they operate over their internal processes), and (ii) model-agnostic techniques, that are compatible with the vast majority of ML models (normally

ICCBR XCBR'22: 4th Workshop on XCBR: Case-based Reasoning for the Explanation of Intelligent Systems at ICCBR-2022, September, 2022, Nancy, France

*Corresponding author.

esteban.brito@aaaimx.org (E. E. Brito-Borges); mauricio.orozco@itmerida.edu.mx (M. G. Orozco-del-Castillo); jareciog@ucm.es (J. A. Recio-García)

0000-0002-4257-944X (E. E. Brito-Borges); 0000-0001-5793-6449 (M. G. Orozco-del-Castillo); 0000-0001-8731-6195 (J. A. Recio-García)

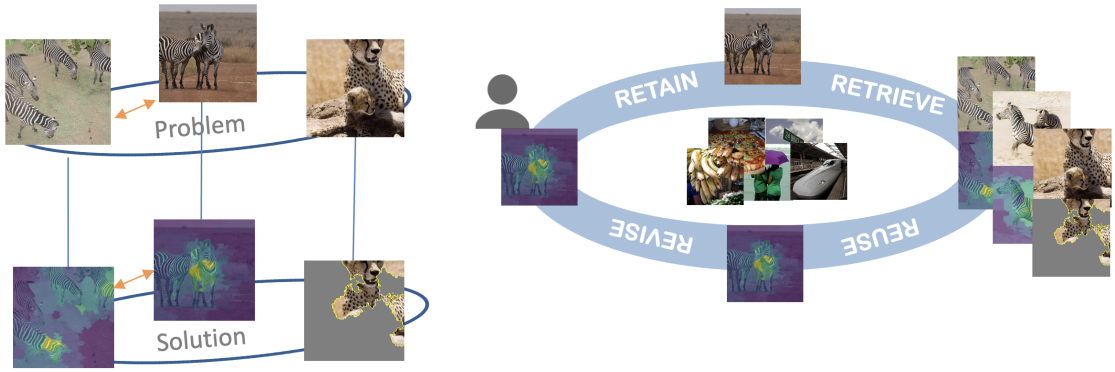

Figure 1: Underlying hypothesis: similar images have similar optimal explanation methods. Left image shows the two dimensions of the cases (description/problem and solution space) following the CBR hypothesis that similar problems have similar solutions. Right side presents a schema of the proposed CBR system.

due to only focusing on perturbation of inputs and not on the internal processes of the models). Examples of model-dependent techniques for the image classification domain are Integrated Gradients (IG) and eXplanation with Ranked Area Integrals (XRAI), while relevant examples of model-agnostic methods include both Locally Interpretable Model-Agnostic Explanations (LIME) and Anchors.

In our previous work [3], these four techniques were analyzed and compared for the task of explaining the predictions given by a deep neural network model trained for image multi-classification. The relative importance of each of these techniques on the task of explaining individual images and their classification was measured through a voting process held with 30 users, providing insight over which of these techniques produced the most humanly-interpretable image explanations. Then, a CBR process was proposed for selecting the most suitable XAI technique for the explanation of new, unseen query images using the experience of previously voted image cases. The conceptual schema or our CBR process is illustrated by Figure 1. Our hypotheses (following the main hypothesis behind CBR) are that: (i) similar images should have the same optimal explanation method given their nature, features, and the classification provided by the DL model; and (ii) the choice of the optimal explanation method has not an algorithmic solution and, therefore, we need to reuse previous explanation experiences to select the most suitable XAI method.

However, in our previous work the computational time required to compute the solution for new query images was proportional to the case base size, i.e., as new cases are added, queries can take more and more time to be solved, damaging the scalability of the process. Therefore, this work proposes the additional hypothesis that clusters of similar images can be aggregated and still have enough information to construct a functional case base for the prediction of most suitable XAI techniques, but reducing considerably the required computational time.

We propose a case base organization using clustering algorithms, where we apply and evaluate several similarity functions, based on different data representations (color histograms, latent features, among others) and distance metrics (Euclidean distance, cosine similarity and structural similarity index).

This paper is organized as follows: Section 2 introduces the XAI algorithms and the clustering techniques. Section 3 describes the clustering-based case elicitation process and the reuse of cases to generate new solutions. In Section 4 we demonstrate the benefits of our approach. Concluding remarks are discussed in Section 5.

2. Background

CBR offers a framework where previously collected experiences can be reused to solve new situations. These situations can incorporate cases where AI systems need to be explained, and so, many initiatives have seen the use of CBR systems to fulfill these tasks, such as the ones discussed in [4].

Recently, an important amount of effort has been assigned to the explanation of black-box models. The majority of these papers perform post-hoc explanations, where CBR is applied after model-agnostic techniques have been used to explain black-box models [5, 6, 7].

Besides the fact that CBR can be used as an explanation method per se, there are other ways to explain how and why AI systems behave the way they do. Several XAI methods have been developed to do this and they work in diverse ways. Model-dependent techniques operate over the internal decision processes of the models they are required to explain. IG and XRAI [8][9] are clear examples of this, where both use backpropagated gradients to fulfill the role of explaining predictions, normally from gradient-based models, like neural networks. Unlike model-dependent methods, model-agnostic techniques like LIME or Anchors [10][11] rely on perturbations of instances in order to accomplish their work of explaining classifiers, being oblivious to the internal logic of the models they explain. This way, the image explanation techniques considered in this paper evenly represent the two main approaches in which a black-box model can be explained: model-dependent and model-agnostic techniques. Figure 2 illustrates the different explanation images generated by these techniques whose details are presented next.

IG In short, the IG technique determines the relationship between the prediction of a black-box model and the features of the instances used as input. To do so, IG is based on the changes of attribution values of every feature with respect to the model's prediction, making use of gradients and partial derivatives [8]. Attributions are calculated over different interpolations between the original instance to be explained and a baseline instance, representing an instance that is completely devoid of information (for example, a black image, if the model processes images). The fact that IG works using gradients limits the range of use of the technique to gradient-based ML models, such as neural networks, although it can be applied over different types of inputs, such as tabular, text-based and visual data.

XRAI On a similar note as IG, XRAI is based on gradients and finds the relationship between the output of a black-box model and the features of those instances used as inputs. Unlike IG, XRAI can only be applied on visual data, because it adds an image segmentation phase right before the attribution calculation phase begins [9]. This produces an additional

advantageous effect, because it is known that segments are preferred over individual pixels when producing explanation images that must be interpreted by humans [12].

LIME LIME a model-agnostic technique that uses perturbations with the goal of explaining the predictions a model generates on the vicinity of an input instance. In that sense, it is also considered a model that produces local explanations, not global ones. LIME does not only focus on generating explanations, but also that these explanations are simple enough in order to facilitate their interpretation.

LIME has compatibility with many types of data: on tabular data, statistical indicators are used over every individual feature, with the goal of generating new perturbations. On image data, LIME groups pixels into segments known as superpixels, and perturbations are generated based on the presence or absence of these regions. Similar behavior is present on text-based data, where the presence or absence of vector words is the main driver behind perturbation generation.

Anchors With the limitations and potential enhancements for LIME in mind, its authors proposed a new technique called Anchors, based mainly on the inclusion of "anchors", sets of if-else conditional rules that, if met completely, guarantee that the prediction assigned to an instances will stay the same [11] (for example, that the same class is predicted with a high level of probability).

Contrary to LIME, Anchors does not rely on a loss function that locally approximates models on the vicinity of an instance, nor does it generates artificial instances based on perturbations. Anchors generates an explanation for instances, but also builds a set A of predicates on which the previous explanation holds for all instances that fulfill the conditions of A. In this sense, Anchors solves the problem of unclear coverage that LIME presented, defining the space where a generated explanation holds.

One of the key points of this paper is the use of clustering to optimize the organization of the case base. Clustering algorithms can be divided into many types, depending on the mechanics used to find groups of similar instances. Particularly, density-based clusters label those areas with a high density of instances as clusters [13]. The exact way density is measured depends on the selected algorithm. On DBSCAN, density is given by two parameters: m and ϵ. An instance is considered as part of a cluster if at least m other instances are within ϵ units of distance from it, or if it is within ϵ units of distance from another instance that fulfills the previous conditions. All other instances are considered as "noise" [14]. The units in which ϵ is defined depend on the metric used to measure similarity between instances.

To identify well-defined clusters, the silhouette score is used. This metric computes the average of silhouette coefficients of every instance used during clustering. This coefficient is defined as:

$$CS_i = \frac{b_i - a_i}{max(a_i, b_i))}, \tag{1}$$

where a_i is the average distance between instance i and all the other instances belonging to the cluster to which it was assigned, and b_i is the average distance between i and all the other

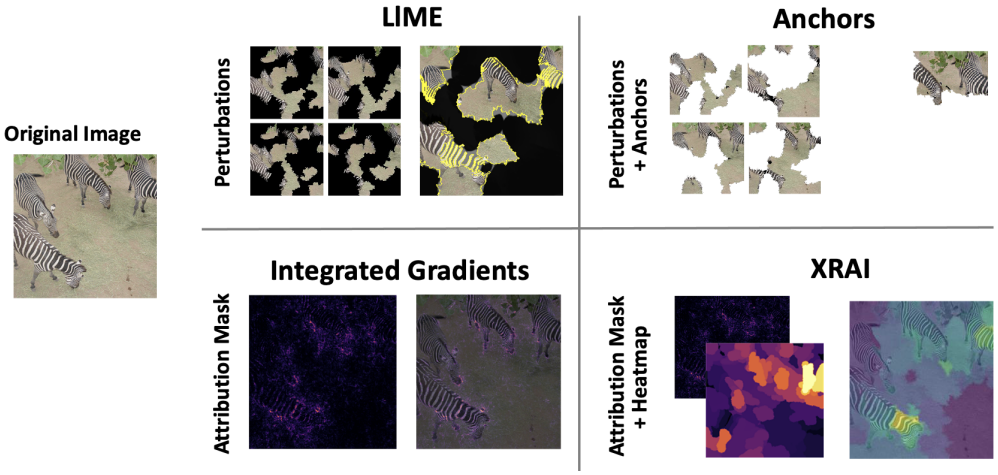

Figure 2: Example of the resulting visual explanations generated by the four techniques considered in this paper.

instances of the nearest cluster to which it was not assigned. The silhouette coefficient (and, by extension, the silhouette score) exist in the range [-1, 1]. A coefficient near +1 indicates that the instance is well placed inside the right cluster, and a coefficient near -1 indicates that it was assigned to the wrong cluster. Coefficients around 0 represent instances that are on the boundaries between clusters [13]. When performing clustering, it is preferred to generate clusters with a silhouette score that approximates to 1.

Next, we present our case base optimization process and the retrieval mechanism used for selecting suitable explanation techniques for new images.

3. Method

As explanations depend on their utility to the user, it is necessary to develop a solution that includes user's opinions on the process used for selecting suitable explanation techniques given an image classification. On a previous paper[3], the use of a CBR approach was proposed where a case base of instances and their corresponding optimal explanation methods are reused to provide an explanation for a given query image. In this paper, concretely, this approach is enhanced implementing clustering techniques that organize the case base before case reuse, greatly reducing the time it takes to generate solutions for new case queries. Next, we present the case base elicitation process, several similarity metrics that have been considered, as well as alternative reuse strategies.

3.1. Case base elicitation

The dataset used in this project consisted of 200 images, extracted from the Visual Genome project [15]. The case base was collected through a voting process where users identify the most suitable techniques for explaining individual pictures and their image classification results (e.g. which technique explains the best why a classifier labeled an image as a dog). This way,

every image is associated to a solution vector, each element of these vectors representing the collected votes for each XAI technique: Anchor, IG, LIME and XRAI.

3.2. Case representation

Every instance in the case base consists of a description and a solution. The description of each case contains three items: the image itself –its pixel matrix M– the color histogram of the image (H), and the latent feature's vector $\vec{f}$ obtained from the internal layers of a DL classifier (in this case, the InceptionV3 convolutional neural network [1]). The solution is the number of votes given by the users to each explanation strategy, denoted as L, A, I and X for LIME, Anchors, IG and XRAI respectively. This representation of cases can be formalized as a description and solution pair $C = \langle D, S \rangle$, where

$$\begin{aligned} D &= \langle M, H, \vec{f} \rangle, \\ S &= \langle L, A, I, X \rangle. \end{aligned} \tag{2}$$

However, under the premise that clusters of similar images can represent the combined information of the individual instances that make part of them, we propose case base optimization using a clustered organization. Lets denote a group of cases $G_p = \{C_1, \ldots, C_i \ldots, C_n\}$ as those cases identified as similar by the clustering algorithm. Then each cluster p is represented by the prototype case $C^p = \langle D^p, S^p \rangle$, where

$$\begin{aligned} D^p &= \langle \mathcal{A}(C_i.M), \mathcal{A}(C_i.H), \mathcal{A}(C_i.\vec{f}) \rangle, \\ S^p &= \langle \overline{C_i.L}, \overline{C_i.A}, \overline{C_i.I}, \overline{C_i.X} \rangle, \\ &\quad \forall\, C_i \in G_p. \end{aligned} \tag{3}$$

Each component of the prototype description can be computed using a custom aggregation function $\mathcal{A}$. In our case, we have chosen the numerical average for each component (pixels, histogram values or features' weights). The solution component of the prototypes is computed through averaging the votes of each technique across the vote solutions from every case inside a cluster G_p.

3.3. Similarity Metrics

As mentioned before, groups of similar images are obtained through a density-based clustering algorithm –DBSCAN– over the dataset. Clustering is based on similarity distances, which express how every instance of a dataset relates to all the others. The same metrics are also used for the retrieval of similar images (and their corresponding vote vectors) for a given query image.

In our work, several distance metrics were used. The *euclidean* and *cosine* distances is only applied over color histograms or latent features, whereas the *Structural Similarity Index* (SSIM) can only be applied on pixel matrices and combines three different factors of an image: luminosity , contrast and structure.

[1] Trained model can be found in the TensorflowHub platform: https://tfhub.dev/google/imagenet/inception_v3/feature_vector/5

Sim. ID	Applied over	Sim. metric
HIST-EUC	Color histogram	Euclidean distance
HIST-COS	Color histogram	Cosine similarity
LF-EUC	Latent features (from InceptionV3)	Euclidean distance
LF-COS	Latent features (from InceptionV3)	Cosine similarity
SSIM	Pixel matrix	Structural similarity

Table 1
Different similarity approaches being considered for the evaluation of the proposed case base clustered organization

Sim. metric	m	ϵ	Sil. Score	#clusters	Clustered instances
HIST-EUC	2	4800	0.57	2	150
HIST-COS	3	0.15	0.45	2	139
LF-EUC	6	0.56	0.56	7	140
LF-COS	6	0.82	0.82	8	150
SSIM	-	-	-	-	-

Table 2
Results obtained from applying clustering to different similarity matrices. SSIM index does not fulfilled the minimum clustering requirements.

Comparison of diverse representations of images using different similarity metrics resulted in multiple configurations of our clustered retrieval method. Those are shown in Table 1, along with identifiers that will enhance the readability of this paper on following sections.

3.4. Clustered case base organization

Several iterations of DBSCAN clustering were performed using every one of the proposed similarity metrics with the goal of generating clusters of similar images. During each clustering iteration, a range of values for DBSCAN parameters (m and ϵ) were tested, and their clustering results compared. For every similarity metric, we selected the values of m and ϵ that resulted in the highest silhouette score while also fulfilling the following restrictions: a minimum of two non-noise clusters were detected in the data and at least 70% of the instances were attributed to some valid cluster (i.e. they were not labeled as "noise" instances). These conditions allowed us to generate well defined clusters using the vast majority of images in our dataset. Clustering parameters and their results are shown in the Table 2. In the case of the SSIM index, no clustering parameters fulfilled the previously mentioned conditions and therefore this similarity metric could not be applied neither evaluated.

3.5. Solution reuse

In our previous work [3], the solution assigned to a new image consisted solely of the most suitable technique for explaining its classification. However, it implies several bias related to closely majoritarian classes. On this paper, the solution of a case is instead a vector that represents the votes for each technique. This way, users are provided with richer information about the relevance of each explanation technique, being able to compare the viability of the four techniques, and picking those that they deem relevant.

As the CBR cycle dictates, solution reuse from similar cases is needed to generate vote predictions. For a new image, its k nearest cases are aggregated to produce this prediction, in the following fashion:

$$predict(S_1, ..., S_k) = \langle L^+, A^+, I^+, X^+ \rangle, \tag{4}$$

where

$$M^+ = \frac{\sum_{i=1}^{k} S_i.M}{k}, \tag{5}$$

Once the CBR process has been defined, following section presents the experimental evaluation of the proposal.

4. Evaluation

In order to evaluate the performance of our process through the clustered organization of the case base, two metrics were defined. The first of them is the Root Mean Squared Error (RMSE) metric, used to calculate distances between the vote vectors estimated by our CBR system and the actual vote vectors assigned to the individual images in our original dataset. The second metric measures the average time (in microseconds) predictions take to be generated. These two metrics help to evaluate 1) the prediction capabilities of our proposed CBR systems and 2) its improvement regarding the computational cost.

RMSE and computational time were calculated using 5-fold cross-validation for both case bases's organizations: the original linear organization where the query is compared to each case (which we'll call instance-based predictions from now onwards), and the new optimized clustered approach.

RMSE values are presented in Figure 3. All the similarity metrics that use individual cases to perform predictions present a high RMSE that quickly converges around a stable, lower value when increasing the k parameter of the k-NN algorithm. For these similarity metrics, the latent features descriptors are the ones that produce the lowest RMSE values. On the other hand, the clustered organization presents an opposite behaviour, growing in error as k does. This happens because the prototype cases encapsulate the information of many instances, and at some point having many prototypes to compared with adds noise and starts to be counterproductive. However, it is clear that all the similarity metrics (except for latent features on instance-based predictions) present RMSE values that do not present any significant advantage over each other. This adds veracity to the hypothesis that clusters of images can be aggregated to create cases that still hold enough information useful to generate predictions.

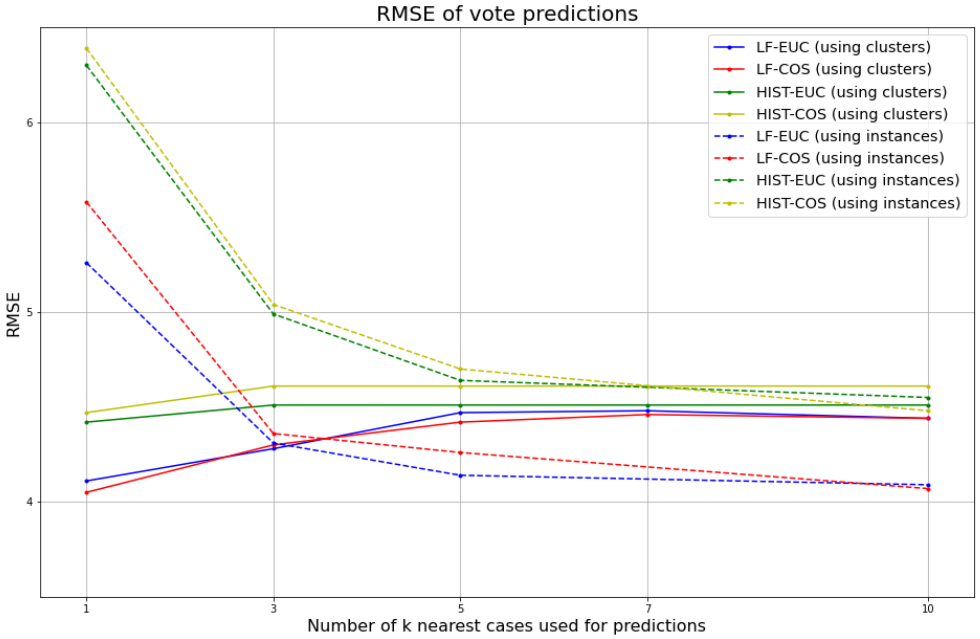

Figure 3: RMSE between vote vector predictions and actual vote vectors. Different similarity metrics and many values for k are being compared. In case of the original linear organization the k value (x-axis) represents the number of top instances returned by the k-NN algorithm. For the clustered organization, the x-axis represents the top prototypes being considered to generate the solution.

When analyzing the average time a prediction takes to be generated for a new query image, significant differences are detected. As shown in Figure 4, all the similarity metrics that use cluster-based vote prediction present a significantly faster query process than instance-based ones. This is due to the organization of the case base, reducing the number of calculations needed to identify the k nearest neighbors for a given query instance. In particular, color histograms present the fastest query phase, possibly attributed to the fact that the dimensionality of their vectors is smaller than those present in latent features.

Results are clear, the clustered organization of the case base obtains an improvement close to 10 times faster than the standard k-NN approach without increasing prediction error significantly.

5. Conclusions

This paper presents a cluster-based approach for the optimization of a case base consisting of images and their corresponding optimal explanation methods according to the users' opinion.

This organization greatly reduces the time required to generate predictions for new queries. The proposed case base optimization process generates groups of cases organized according their estimated optimal explanation method. Our approach applies and evaluates the DBSCAN clustering method over different similarity metrics.

To evaluate our proposal, a K-fold cross-validation process was performed, both on the clustered organization and the original case base. Predictions obtained from clusters computed

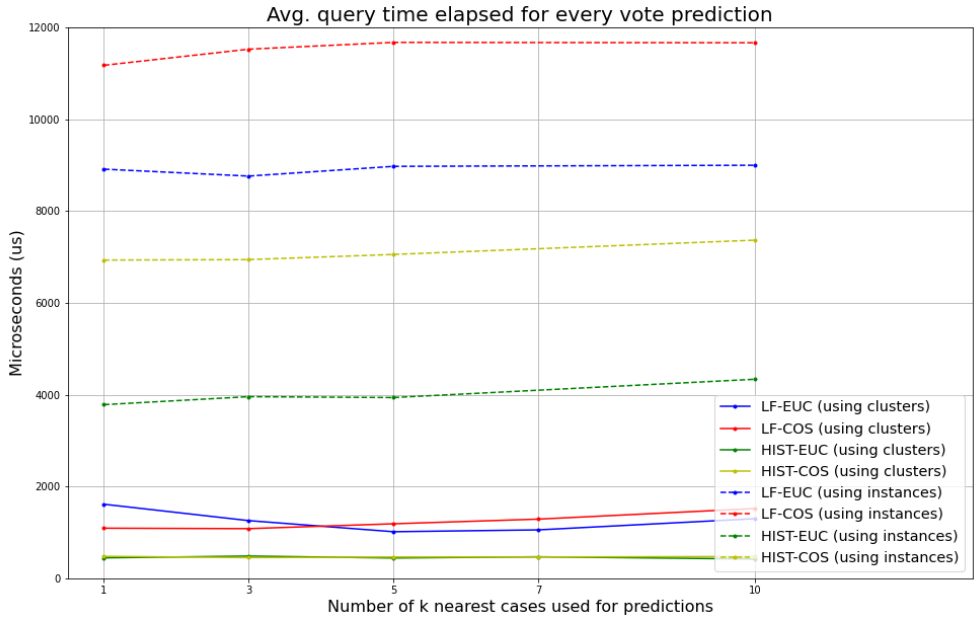

Figure 4: Average query times for different similarity metrics and several values for k. In case of the original linear organization the k value (x-axis) represents the number of top instances returned by the k-NN algorithm. For the clustered organization, the x-axis represents the top prototypes being considered to generate the solution.

using any similarity metric applied over the latent features of the image classifications, present the lowest error rate, quantified using the RMSE metric, while histogram-based metrics resulted to have the fastest prediction times for new query images.

From this evaluation, it can be concluded that, indeed, clusters of similar images can be aggregated to reduce the complexity of the case base and still have enough information to be used in the accurate prediction of suitable XAI techniques. With the added benefit of being up to 10 times faster when processing new cases.

Many improvements can be made to enhance the performance, coverage and validity of this work. First of all, our current study only compared four explanation techniques, but many more are being presented to the XAI community, such as the ones proposed by [16]. Also, it is known that image class imbalance was present on the case base taken from [3], so a evenly distribution of classes would help achieve more confidence on the proposed CBR processes.

Other image representation features can be used during clustering, incorporating more perspectives than just color histograms or latent features. These features can be enhanced too, selecting vectors produced in more advanced deep learning classification models, such as InceptionV4, or even other types of CNN implementations.

Finally, although results proved that a case base organization process through aggregation of clustered cases achieves relevant prediction performance, it is not clearly known how or why the use of clustering causes this, and so, further research is needed in order to make this process self-explainable.

Acknowledgments

This research is a result of the Horizon 2020 Future and Emerging Technologies (FET) programme of the European Union through the iSee project (CHIST-ERA-19-XAI-008, PCI2020-120720-2) funded by MCIN/AEI/10.13039/501100011033 and European Union "NextGenerationEU"/PRTR. This work has been also partially supported by the PERXAI project PID2020-114596RB-C21 funded by MCIN/AEI/10.13039/501100011033 and the BOSCH-UCM Honorary Chair on Artificial Intelligence applied to Internet of Things of the Universidad Complutense de Madrid.

References

[1] P. Hall, N. Gill, An introduction to machine learning interpretability : an applied perspective on fairness, accountability, transparency, and explainable AI, 1 ed., O'Reilly Media, Inc., Sebastopol, 2018. URL: https://www.oreilly.com/library/view/an-introduction-to/9781492033158/.

[2] R. Mac, Facebook Apologizes After A.I. Puts 'Primates' Label on Video of Black Men, 2021. URL: https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html.

[3] J. A. Recio-García, H. Parejas-Llanovarced, M. G. Orozco-del-Castillo, E. E. Brito-Borges, A case-based approach for the selection of explanation algorithms in image classification, in: Case-Based Reasoning Research and Development, ICCBR 2021, volume 12877 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 186–200. doi:10.1007/978-3-030-86957-1_13.

[4] D. Leake, D. McSherry, Introduction to the Special Issue on Explanation in Case-Based Reasoning, Artificial Intelligence Review 24 (2005) 103–108. URL: https://link.springer.com/article/10.1007/s10462-005-4606-8. doi:10.1007/S10462-005-4606-8.

[5] M. T. Keane, E. M. Kenny, How Case-Based Reasoning Explains Neural Networks: A Theoretical Analysis of XAI Using Post-Hoc Explanation-by-Example from a Survey of ANN-CBR Twin-Systems, in: ICCBR 2019: Case-Based Reasoning Research and Development, volume 11680 LNAI, Springer Verlag, 2019, pp. 155–171. doi:10.1007/978-3-030-29249-2_11/COVER/.

[6] O. Li, H. Liu, C. Chen, C. Rudin, Deep Learning for Case-Based Reasoning Through Prototypes: A Neural Network That Explains Its Predictions, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018. doi:10.1609/AAAI.V32I1.11771.

[7] L. Gates, C. Kisby, D. Leake, CBR Confidence as a Basis for Confidence in Black Box Systems, in: ICCBR 2019: Case-Based Reasoning Research and Development, volume 11680 LNAI, Springer Verlag, 2019, pp. 95–109. doi:10.1007/978-3-030-29249-2_7.

[8] M. Sundararajan, A. Taly, Q. Yan, Axiomatic Attribution for Deep Networks, in: 34th International Conference on Machine Learning, Sydney, 2017. URL: http://proceedings.mlr.press/v70/sundararajan17a.html.

[9] A. Kapishnikov, T. Bolukbasi, F. Viegas, M. Terry, XRAI: Better attributions through regions, Proceedings of the IEEE International Conference on Computer Vision 2019-October (2019) 4947–4956. doi:10.1109/ICCV.2019.00505. arXiv:1906.02825.

[10] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?" Explaining the predictions of any classifier, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 1135–1144. doi:10.1145/2939672.2939778.

[11] M. T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-Precision Model-Agnostic Explanations, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018, pp. 1572–1535. URL: https://ojs.aaai.org/index.php/AAAI/article/download/11491/11350.

[12] M. Sundararajan, S. Xu, A. Taly, R. Sayres, A. Najmi, Exploring Principled Visualizations for Deep Network Attributions, in: Joint Proceedings of the ACM IUI 2019 Workshops, Lon Angeles, 2019.

[13] A. Géron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2 ed., O'Reilly Media, 2019.

[14] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, in: E. Simoudis, J. Han, U. Fayyad (Eds.), Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD96), AAAI, Portland, 1996.

[15] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, M. S. Bernstein, L. Fei-Fei, Visual genome: Connecting language and vision using crowdsourced dense image annotations, Int. J. Comput. Vision 123 (2017) 32–73. doi:10.1007/s11263-016-0981-7.

[16] S. Xu, S. Venugopalan, M. Sundararajan, Attribution in scale and space, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, USA, 2020, pp. 9677–9686. doi:10.1109/CVPR42600.2020.00970. arXiv:2004.03383.