Pseudo Labeling and Classification of High-Dimensional Data using Visual Analytics

Bárbara C. Benato*
University of Campinas, Brazil
Utrecht University, the Netherlands

Alexandru C. Telea
Utrecht University, the Netherlands

Alexandre X. Falcão
University of Campinas

***Abstract*—Machine learning (ML) works with data consisting of tens up to tens of thousands of measurements (dimensions) per sample. As the number of dimensions and/or samples grow, so does the difficulty of understanding such data and its ML pipelines. Visualization, and in particular Visual Analytics (VA) has emerged as one of the key approaches that helps practitioners with the understanding of high-dimensional data and with ML engineering tasks. In this paper, we investigate several novel approaches by which VA can help ML (and conversely). Our work focuses on a visualization technique called *dimensionality reduction*, or projection, and the task of training a classifier when only a small amount of ground-truth labels is available. As result, experiments show that projections can capture very well the data structure present in high dimensions to support the design of high-performance feature and classifier learning models. Also, experiments relate projection quality to data separation and classifier performance. Finally, we combine these two observations to assist users in manual labeling samples to show that both algorithms and humans can exploit projections to build better classifiers. We argue that the ability of pseudo labels in retain information from 2D projected spaces is the key idea that links all these contributions.**

I. Introduction

Machine learning (ML) works, in most cases, with *high dimensional data*. By this, we mean datasets consisting of samples (data points) which have, each, tens up to thousands of different measurements (dimensions). As the size – either in number of rows (samples) or columns (dimensions) – of these data tables grows, so does the difficulty of *understanding* them and, even more importantly, understanding how ML pipelines process them.

For the last several decades, data visualization (VIS) has grown aside and along machine learning (ML). In its early phases, visualization has been introduced to science and engineering fields by the need of understanding increasingly large (and complex) datasets generated either by measurements or by numerical simulations produced by scientific computing applications. However, even more importantly, all data visualization applications share the same ultimate goal – to enable their users to gain so-called *actionable insights* into the phenomena which have generated the explored data, and, next, to use such insights to improve other aspects of the processes involved with the respective data [1], [2]. Recognizing this commonality between all visualization applications, a new field called *visual analytics* (VA) has emerged.

As such, it is not surprising that visualization (VIS) and visual analytics (VA), with their stated aims of helping users extract actionable insights from large and complex datasets, has emerged as a key tool in assisting ML engineering. Among the family of visualizations techniques that handle such high-dimensional data, *dimensionality reduction* (DR) techniques, also called projections, have emerged as one of the most successful and most frequently used to assist ML engineering. As mentioned above, datasets in ML pipelines are, by their very nature, high dimensional. From such datasets, DR techniques create typically 2D scatterplots, one plot point per data sample. In this process, DR techniques aim to preserve the so-called *data structure*. That is, characteristics of the input *dataset* which are deemed important for the problem at hand should be encoded by similar characteristics measured by the corresponding scatterplot points, which may be used to reason about the input data.

Apart from the above-mentioned added value of high-dimensional visualizations (in particular, projections) for ML engineering, the opposite interaction exists too. That is, ML can be used to create better visualizations for high-dimensional data. From these, our general research question can be stated quite succinctly as follows:

- **RQ)** How can we exploit the synergy between ML and DR techniques to improve each other?

It is, however, evident that the scope and generality of the above-stated research question are too high to be able to fully answer it. As such, we further refine this general research question as follows:

- **RQ1)** How to use multidimensional projections to build better models for machine learning?
- **RQ2)** How does projection quality relate to data separation and classification performance?

RQ1 relates to the observation that DR projections of high quality capture well many aspects of the so-called data structure, *i.e.*, the relative organization of data samples in a high-dimensional dataset. If and when this is the case, it means that the information that a projection captures can be used as a 'proxy', or substitute, to the original information that the high-dimensional samples themselves captures. In turn, this means that one can use the projection *instead of*, and in places where,

*This paper is related to the thesis of Bárbara C. Benato resulting from a joint Ph.D. program between the University of Campinas (Brazil) and Utrecht University (the Netherlands).

one would use the high-dimensional information, and obtain, ideally, very similar – if not better – results.

RQ2 relates to the same observation of RQ1. Let us assume that a projection captures well data structure by means of the visual separation (VS) it exhibits in its scatterplot. Separately, it is well known that the performance of a trained ML classification model (classifier performance, or CP) relates to the data separation (DS) present in the dataset it works on. Simply put, it is easier to classify a dataset with high DS to reach a high CP than a dataset with a low DS. This means that all three above-mentioned factors – DS, VS, and CP – are inter-related. If such a relation is true, we can exploit such observations to *measure* and/or *optimize* one of these factors as function of the other two.

The paper is organized as follows. Section II addresses RQ1 by using DR techniques to extract useful features from a high-dimensional dataset for creating an end-to-end ML pipeline for a challenging classification problem. Section III focuses on RQ2 by extending earlier observed relations between DS and VS, respectively between DS and CP, and showing how one can pre-process high-dimensional data to achieve a high DS, which in turn leads to a high VS in projections of that data; and finally, using such projections to achieve a high CP. Section IV exploits the results presented in Sec. III to further answer RQ2. Specifically, if high-VS projections can be used to construct high-CP models, we can use existing ML quality metrics to measure the visual separation (VS) present in a given projection. Section V revisits RQ1 by studying an added-value of projections for helping the construction of high-performance ML models. Specifically, we now consider the user-in-the-loop scenario by asking subjects to use an interactive projection-based tool to create pseudo labels so as to train a classification model. Finally, Section VI revisits our contributions to solving RQ1 and RQ2 and also outlining open avenues for future research.

II. Feature and Classifier Learning

The success of deep neural networks (DNNs) is evident in many applications. A critical problem appears when the training set is too small [3] – *e.g.* only dozens of supervised samples per class. We propose a meta-pseudo-labeling methodology, called *Deep Feature Annotation* (*DeepFA*) [4]–[6], to train DNNs from very few supervised samples (e.g., 1% of a dataset) without a validation set. In *DeepFA*, the teacher – a connectivity-based semi-supervised classifier, Optimum Path Forest (OPFSemi) [7]– exploits modifications of a given latent feature space of the student (a DNN) along with iterations of non-linear 2D projection (t-SNE [8]) for pseudo-labeling. At each iteration, the most confidently labeled samples are used to retrain the DNN, modifying its latent feature space. The semi-supervised classifier does not require parameter optimization, dismissing a validation set (extra labeled samples). The number of labeled samples is next increased to improve the DNN with pseudo-labeled training set.

Isolated aspects of *DeepFA* using OPFSemi on a 2D embedded space have been evaluated. In [4], a few iterations of the training loop with truly-and-artificially-labeled samples was enough to improve the generalization performance of a supervised DNN. We call this version as *orig-DeepFA*, while we use *DeepFA* to refer to the entire methodology. OPFSemi's confidence was also considered when selecting unsupervised samples to train the supervised DNN [5], reducing label propagation errors – called as *conf-DeepFA*.

However, *conf-DeepFA*'s performance on test sets can oscillate along with the pseudo-labeling iterations such that the model obtained at the last iteration is not guaranteed to be the best model. To circumvent this problem, we propose *ext-DeepFA*, which extends *conf-DeepFA* by computing a clustering-based metric from the pseudo-labeled samples to select the optimal model for generalization among the ones generated along with the pseudo-labeling iterations. While earlier *orig-DeepFA* and *conf-DeepFA* variants have shown promising results, the methods may differ in the deep architecture used for feature learning and classification, the semi-supervised classifier for label propagation, the projection technique, and the criterion to select the model for generalization. The present work is then a comprehensive study on DeepFA.

A. Experimental evaluation

We next outline how we organized our study.

1) Datasets: We choose eight diverse datasets to perform our investigations, as follows.

MNIST: We first chose the public MNIST [9] dataset, to explore a known and easy classification task. MNIST has 0 to 9 handwritten digits grayscale images (28×28 pixels). We use 5000 random samples from the original training dataset.

Parasites: The next five datasets come from a Parasite medical image collection [10]. This collection has three main dataset types: (i) *Helminth larvae*, (ii) *Helminth eggs*, and (iii) *Protozoan cysts*. The datasets contain color microscopy images (200×200 pixels) of the most common species of human intestinal parasites in Brazil, responsible for public health problems in most tropical countries [10]. The datasets are challenging since they are unbalanced and contain a majority impurity class, with samples very similar to parasites, making classification hard (see Fig. 1). To these datasets, we add the (iv) *Helminth eggs* and (v) *Protozoan cysts* without the impurity class datasets, yielding a total of 5 datasets.

Coconut: We use a random subset of the Coconut trees dataset [11] with $7,827$ regions (90×90 pixels) of aerial colored images from the Kingdom of Tonga, acquired by satellite imagery in October 2017, labeled by Humanitarian OpenStreetMap. The dataset has two classes: images with $(6,139)$ or without coconut trees $(1,688)$.

COVID: A team of researchers from the Universities of Qatar and Dhaka have created a database [12], [13] of chest X-ray images (299×299 pixels). We obtained the second update of the dataset with 21165 images split in four classes: COVID-19 positive (3616), lung opacity (non-COVID lung infections, 6012), viral pneumonia (1345), and normal (10192) cases. We use a randomly subset of this dataset with 10583 images.

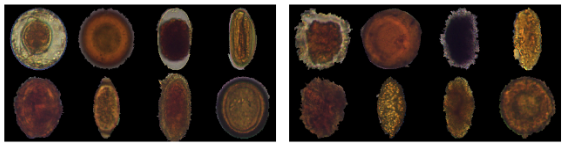


Fig. 1. Examples of *H.Eggs* species (left) and similar images of impurities (right).

B. Experimental setup

To reproduce the scenario of few supervised samples, we define a supervised training set S with only 1% of supervised samples from a given dataset D. The unsupervised U and test T sets have 69% and 30% of samples, respectively ($D = S \cup U \cup T$). The small S simulates the real-world scenario when one has a large D but manual effort is needed to label samples to create S. We randomly divide each dataset D into S, U, and T in a stratified manner and also generate three distinct splits for each experiment for statistical analysis. We evaluate our method by the probability of the chosen deep architecture's last fully-connected layer, *i.e.*, just before the classification layer. From this, we compute accuracy and κ. We evaluate label propagation accuracy by computing the number of correctly assigned labels in U.

C. Experimental results

1) Different pseudo labeling strategies: Using different pseudo-labeling methods within the *orig-DeepFA* looping means that the label propagation and the learned feature space can be mutually affected. To evaluate how, Fig. 2 shows the resulting feature space and label estimation of the two best methods found in evaluated experiments – which can be found in the thesis –, *i.e.*, *OPFSemi* and *L.Spread*$_{rbf}$. In this figure, we use a dataset with and distinct classification values (H.larvae, $\kappa \in \{0.80, 0.06\}$).

For H.larvae, *OPFSemi* was able to propagate labels only for regions with supervised samples of class 1 (red) and also provide a feature space (2D projection) in which class 1 is separated from other samples in the projection. In contrast, *L.Spread*$_{rbf}$ was more conservative in propagating labels for class 2 (green) *vs* class 1 (red) so that only the closest samples of class 2 were labeled with that class. At a higher level, Fig. 2 illustrates how distinct ways of propagating labels intervene in the learned feature space produced by the proposed *orig-DeepFA* looping.

2) Feature space improvement over iterations: Figure 3 shows the plot for train and validation loss and accuracy considering 20% (from S) as validation set during one split of MNIST training. The initial learning curve and the learning curves for each iteration are also shown. The learning curves show that the labeled samples can improve the network convergence along iterations. As future work, a different deep network can be tested at the final stage. Also, an unsupervised quality measure can be proposed to define the best feature space found at certain iterations and, hence, the best iteration of the method.

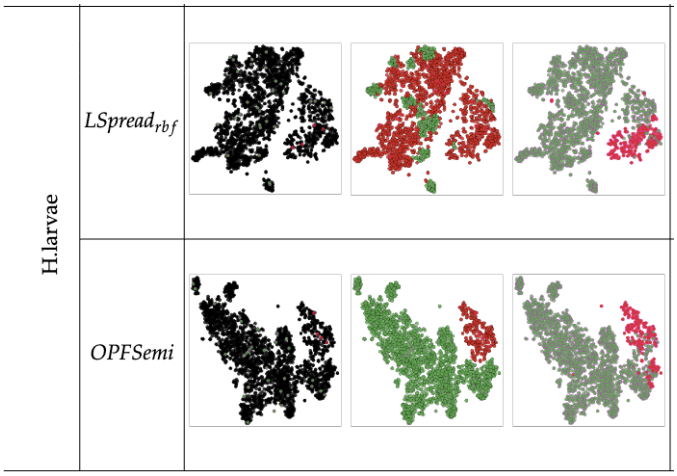


Fig. 2. Comparison of *DeepFA* using *LSpread*$_{rbf}$ and *OPFSemi* pseudo labeling for H.Larvae datasets, with 1% supervised samples and last iteration out of five. 2D feature-space projections of training samples ($S \cup U$) in columns per dataset (from left to right): supervised samples colored by true labels (red=1, green=2), unsupervised ones are black; samples colored by assigned pseudo labels; and samples colored by their true labels. Classification results (per class and total) are shown on the right.

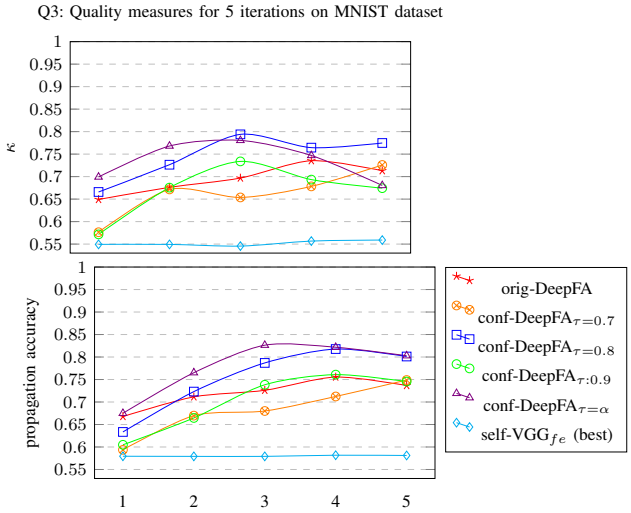


Fig. 3. Results of κ (top) and propagation accuracies (bottom) for the MNIST dataset in one split over 5 iterations, considering self-VGG-16$_{fe}$ (best result), *orig-DeepFA*, and *conf-DeepFA* experiments.

D. Discussion

Our findings confirm that the t-SNE projection technique – well known in information visualization for its high ability to capture *data structure* present in high dimensions – can generate 2D feature spaces which are also effective for feature learning and classifier engineering. In other words, we have shown that 2D projections can be used instead of the original feature space, which typically has hundreds of dimensions or even more, for designing feature and classifier learning models through pseudo-labeling.

III. Linking Data Separation, Visual Separation, and Classifier Performance

Both pseudo-labeling, and broader, the success of training a classifier, depend on a key aspect – how easy is the data *separable* into different groups of similar points. Projections, or dimensionality reduction methods, are well known techniques that aim to achieve precisely this [14], [15]. The success of such tasks involving projections depends on the *visual separation* (VS) of the projection used to depict it. If a dataset exhibits clear *data separation* (DS) into samples of different classes, then analysts should be able to gauge this by seeing a corresponding *visual* separation in the projection, in terms of densely-packed, ideally non-overlapping, groups of points with the same label (within a given group). Conversely, if a dataset exhibits poor data separation, its projection should also show poor visual separation. Although projections have been used to explain and relate many machine learning (ML) concepts, to our knowledge, no work so far has explored the relationship between data DS, VS, *and* classifier performance (CP) in the context of using pseudo-labeling.

We address the above by studying how to generate a high DS using *contrastive learning*. We evaluate DS by measuring CP for a classifier trained with only 1% supervised samples. Then, we evaluate VS fed with the encoder's output of our trained contrastive models. Lastly, we investigate CP by using our above pseudo-labeling to train a deep neural network. We perform all our experiments in the context of a challenging medical application (classifying human intestinal parasites in microscopy images). Our main contributions are as follows:

- C1: We use contrastive learning to reach high DS;
- C2: We show that projections constructed from contrastive learning methods (with good DS) lead to a good VS between different classes;
- C3: We train classifiers with pseudo-labels generated via good-VS projections to achieve a high CP;
- C4 We identify projection techniques for which DS strongly correlates with VS and also techniques for which this does not happen;
- C5 We show that good-VS projections are essential for training classifiers that reach a high CP.

Jointly taken, our work brings more evidence that VS, DS, and CP are strongly correlated and that this correlation, and 2D projections of high-dimensional data, can be effectively *used* to build higher-CP classifiers for the challenging case of training-sets having very few supervised (labeled) points.

IV. Measuring Visual Separation in Projections

Many projection methods have been proposed, using different underlying techniques as graphs, linear algebra, optimization, and neural networks [14]. Such techniques generate a wide variety of scatterplots for the same give dataset, especially when one changes their various hyperparameters. Several *metrics* have been proposed to quantify a projection's quality. However, the most used metrics in the DR literature – Trustworthiness (T) [16], Continuity (C) [16], Normalized stress (S) [17], and Neighborhood hit (N) [18] do not directly measure visual separation at a global projection level but rather more local properties.

We proposed a new VS quality assessment approach based on ML techniques. We exploit earlier findings that studied VS in t-SNE projections to propagate labels, also called pseudo labeling. Projections with high VS (as assessed qualitatively by users) led to good label propagation results [19]. Our hypothesis is that the converse is also true: If we measure a good label propagation score, then the projection will have a high VS. For label propagation, we use the semi-supervised optimum path forest algorithm (OPFSemi) [7] in the 2D projection space provided by DR methods. OPFSemi was shown to lead to very good label propagation accuracies in both high-dimensional and low-dimensional spaces [6], [7] and as such is a good candidate for this task. We evaluate the label propagation by computing the coefficient of agreement of Cohen's Kappa (κ) [20] between true and pseudo labels, a simple but fast and effective way to perform this task which works well also for unbalanced labeled datasets.

Fig. 4 shows our VS measurement pipeline which is detailed next.

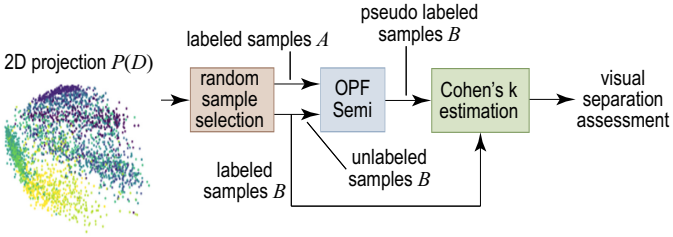


Fig. 4. Pipeline of our approach to assess VS in projections.

We assess our proposal on 39 projection algorithms for 18 labeled datasets and show that our method correlates with perceived VS (measured by a user study) better than well-known metrics for projection quality used in the DR literature. As such, we argue that our metric is an additional useful way to characterize the quality of a projection, atop of existing projection quality metrics.

V. Active Learning using Decision Boundary Maps

Our work so far has shown the added value of projection techniques for the generation of pseudo labels for constructing high-performance classification models. All in all, the above show that projections, and label propagation algorithms in projection spaces, are useful instruments for *automating* both ML and VA tasks. However, this potentially conveys the incorrect message that the *user* has no actual place in such pipelines. We argue that this is by far not the case. The quality of the pseudo labels produced in [21] was constrained by the initial feature space. If the feature learning step produces a feature space with poor visual separation, then an automatic and/or manual label propagation techniques

would fail. We circumvented this problem in Section II by proposing to improve the feature learning over iterations of pseudo labeling using 2D projections. However, this approach left the user (manual annotation) out of the loop. Figure 5 clarifies the above by comparing the pipeline we have used in our previous DeepFA workflows with the interactive pipeline that uses Decision Boundary Maps (DBMs) [22] and error maps (further described in this chapter).

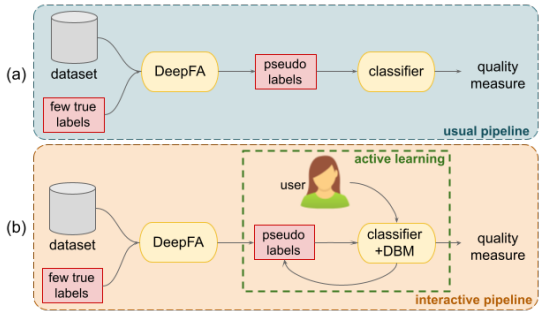


Fig. 5. (a) Generic pipeling using DeepFA. (b) Interactive pipeline in an active learning scenario using DeepFA and VA tools.

We next aim to complete our quest for assessing the added-value of manual labeling by incorporating more advanced VA techniques in support of this user task. Specifically, we consider using DBMs and direct-and-inverse projection errors as visual aids to help users decide where, in a projection, they should concentrate their manual labeling efforts. If the user is able to get insights about the classifier's decision, then the user can (arguably) successfully intervene in its training by manually propagating labels as a way to leverage both propagation and classifier performance. We achieve this by adding an active learning looping in the classifier step of our proposed DeepFA pipeline (Section II). We next describe in detail our proposed pipeline that combines VA techniques and manual labeling in an active learning looping (see also Figure 6).

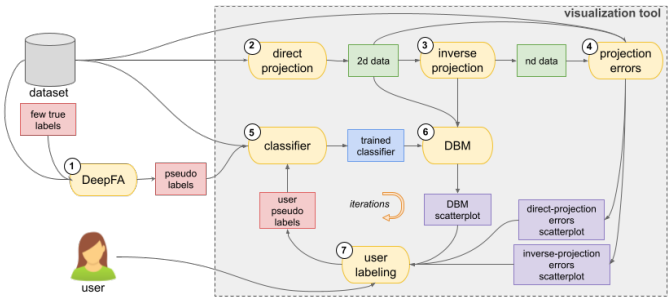


Fig. 6. Pipeline of the proposed interactive approach.

1) Visualization plot: Figure 7 shows a visualization of the combined projection error ϵ computed using the inverse projection technique described above. Error values are encoded into brightness (high ϵ are bright; low ϵ are dark, pixels, respectively). As visible, the projection errors are low close to most of the actual scatterplot points, which is expected, since the projection technique used here (t-SNE) is known to have low errors everywhere on the considered dataset (MNIST). As we go further from the projected points, we see how errors increase.

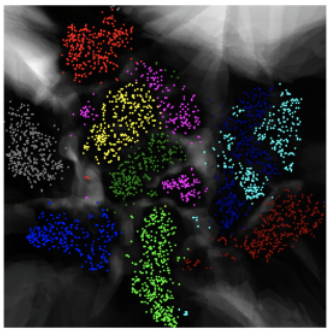
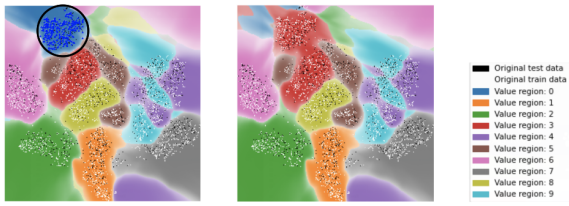


Fig. 7. Scatterplot image with 2D projected points for the MNIST dataset with points colored by class. Projection errors ϵ are computed using the inverse projection technique at every image pixel. Light regions refer to a high error values; dark regions represent low error values.

2) User interaction: Figure 8 shows scatterplots (a) before and (b) after a manual labeling iteration, including the subsequent classifier re-training and DBM recalculation. In this example, for illustration purposes, the user manually selected a large set of points in the blue decision zone (marked by the black circle) and assigned them the label 3 (red). Image (b) shows how the re-trained classifier now has a large red decision zone that includes largely all the points the user has manually labeled as red. However, due to this massive re-labeling, the classifier's performance decreases significantly – accuracy drops from 0.8793 to 0.7927; κ drops from 0.8685 to 0.7689. This is, of course, expected, given that the user has basically forced the disappearance of roughly the entire blue decision zone. In practical use, manual labeling will select significantly fewer samples to label during an iteration.



(a) (b)

Fig. 8. Example of classifier re-training and DBM re-calculation. (a) Initial state (classifier accuracy: 0.8793, κ: 0.8685). User selects a large set of points in the blue decision zone (marked by the black circle) and decides to manually assign them the label 3 (red). (b) Situation after classifier re-training with the new manually added labels (classifier accuracy: 0.7927; κ: 0.7689).

If the classifier performance decreased (as shown in the tool's interface) as compared to the previous iteration, the user can decide to *undo* the last-performed labeling. The labeling window then changes to show the values (DBM, classifier performance, κ) before this past iteration. The process continues until the user decides to stop it, either because of time

constraints or because the desired classifier performance has been reached.

VI. Conclusion

We close this paper by revisiting our proposed contributions. We addressed RQ1 in Section II by proposing a pseudo-labeling approach, called DeepFA, that explores the ability of a multidimensional projection to generate a reduced (two-dimensional) feature space with enough information to improve feature learning and classifier performance over iterations. We have shown that 2D projections can be used instead of the original feature space for designing feature and classifier learning models through pseudo-labeling.

RQ2 is explored in Section III by exploring pseudo-labeling and projection techniques to link DS, VS, and CP. We demonstrated the correlation between DS, VS, and CP, and indicated which *specific* projection techniques preserve a strong DS-VS-CP correlation (or not). Knowing which techniques exhibit this property is valuable for supporting classifier engineering and, arguably, for other infovis applications where VS is important.

We further address RQ2 in Section IV by proposing to assess the VS of 2D projections by a metric that evaluates the CP of a graph-based semi-supervised classifier used to propagate labels in the 2D projection. We claim that, when a classifier achieves high CP in the labeling task, then the 2D projection has a good VS. Our findings show that our proposed metric can better gauge VS in projections than projection-quality metrics commonly used in the DR literature.

Finally, in Section V, we turn back to RQ1. We considered assisting the human *user* to improve the construction of classification models (by means of automatic pseudo labeling). At a higher level, the work in this chapter answers RQ1 by showing that *both* humans and machines can use multidimensional projections to build better classifiers; and that, when *cooperating* (by using such projections), humans and machines obtain results which surpass what can be obtained when using only manual, or only automatic, methods.

Dimensionality reduction techniques and pseudo-labeling play an essential role in all our contributions. Dimensionality reduction successfully captures data structure from high dimensions in a way that machines and humans can distill relevant information to design and improve machine learning models. Pseudo labeling can retain, and next create, substantial knowledge in the produced labels during the entire data flow from the learned space, to the reduced 2D projection space, and next to the way users abstract this information during their iterative manual exploration and labeling of data via 2D projection spaces.

Acknowledgment

The authors acknowledge CAPES (# STIC-AMSUD 88887.878869/2023-00), CNPq (#304711/2023-3, #441950/2923-3), USP CEPID-CeMEAI (#2013/07375-0), FAPESP (#2023/14427-8, #2019/10705-8, #2022/12668-5).

Publications

The thesis is the result of the following publications (in chronological order): [4], [5], [6], [19], [23], [24], [25]. Other publications during the development of the thesis include: [26]–[29].

References

[1] N. Andrienko, G. Andrienko, G. Fuchs, A. Slingsby, C. Turkay, and S. Wrobel, *Visual Analytics for Data Scientists*. Springer, 2020.

[2] K. A. Cook and J. J. Thomas, "Illuminating the path: The research and development agenda for visual analytics," Pacific Northwest National Laboratory (PNNL), Richland, WA (US), Tech. Rep., 2005.

[3] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era." in *Proc. ICCV*, 2017, pp. 843–852.

[4] B. C. Benato, J. F. Gomes, A. C. Telea, and A. X. Falcão, "Semi-supervised deep learning based on label propagation in a 2D embedded space," in *Proc. CIARP*. Springer, 2021, pp. 371–381.

[5] B. C. Benato, A. C. Telea, and A. X. Falcao, "Iterative pseudo-labeling with deep feature annotation and confidence-based sampling," in *Proc. SIBGRAPI*. IEEE, 2021, pp. 192–198.

[6] B. C. Benato, A. C. Telea, and A. X. Falcão, "Deep feature annotation by iterative meta-pseudo-labeling on 2d projections," *Pattern Recognition*, vol. 141, p. 109649, 2023.

[7] W. Amorim, A. Falcão, J. Papa, and M. Carvalho, "Improving semi-supervised learning through optimum connectivity," *Pattern Recognit.*, vol. 60, pp. 72–85, 2016.

[8] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, 2014.

[9] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010, http://yann.lecun.com/exdb/mnist.

[10] C. Suzuki, J. Gomes, A. Falcão, S. Shimizu, and J.Papa, "Automated diagnosis of human intestinal parasites using optical microscopy images," in *Proc. Symp. Biomedical Imaging*, April 2013, pp. 460–463.

[11] J. E. Vargas-Muñoz, P. Zhou, A. X. Falcão, and D. Tuia, "Interactive coconut tree annotation using feature space projections," in *Proc. IGARSS*, 2019, pp. 5718–5721.

[12] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al Emadi *et al.*, "Can ai help in screening viral and covid-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020.

[13] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan *et al.*, "Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images," *Comput. Biol. Med.*, vol. 132, p. 104319, 2021.

[14] L. Nonato and M. Aupetit, "Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment," *IEEE Trans. Vis. Comput. Graph*, 2018.

[15] M. Espadoto, R. Martins, A. Kerren, N. Hirata, and A. Telea, "Toward a quantitative survey of dimension reduction techniques," *IEEE TVC*, vol. 27, no. 3, pp. 2153–2173, 2019.

[16] J. Venna and S. Kaski, "Visualizing gene interaction graphs with local multidimensional scaling," in *Proc. ESANN*, vol. 6, 2006, pp. 557–562.

[17] P. Joia, D. Coimbra, J. A. Cuminato, F. V. Paulovich, and L. G. Nonato, "Local affine multidimensional projection," in *Proc. IEEE TVCG*, 2011, pp. 2563–2571.

[18] F. V. Paulovich, L. G. Nonato, R. Minghim, and H. Levkowitz, "Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping," *IEEE TVCG*, pp. 564–575, 2008.

[19] B. C. Benato, A. X. Falcão, and A.-C. Telea, "Linking data separation, visual separation, and classifier performance using pseudo-labeling by contrastive learning," in *Proc. VISAPP*, 2023.

[20] J. L. Fleiss and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability," *Educ. Psychol. Meas.*, vol. 33, no. 3, pp. 613–619, 1973.

[21] B. C. Benato, J. F. Gomes, A. C. Telea, and A. X. Falcão, "Semi-automatic data annotation guided by feature space projection," *Pattern Recognit.*, vol. 109, p. 107612, 2021.

[22] M. F. C. Rodrigues, R. Hirata, and A. Telea, "Image-based visualization of classifier decision boundaries," in *Proc. SIBGRAPI*, 2018, pp. 353–360.
[23] B. C. Benato, A. X. Falcao, and A. C. Telea, "Linking data separation, visual separation, and classifier performance using pseudo-labeling by contrastive learning," in *Proc. VISAPP*. SciTePress, 2023.
[24] B. C. Benato, A. X. Falcão, and A. C. Telea, "Measuring the quality of projections of high-dimensional labeled data," *Computers & Graphics*, vol. 116, pp. 287–297, 2023.
[25] B. C. Benato, C. Grosu, A. X. Falcao, and A. C. Telea, "Human-in-the-loop: Using classifier decision boundary maps to improve pseudo labels," in *submitted*, 2024.
[26] M. Roder, L. A. Passos, L. C. F. Ribeiro, B. C. Benato, A. X. Falcão, and J. P. Papa, "Intestinal parasites classification using deep belief networks," in *International Conference on Artificial Intelligence and Soft Computing. Springer*, 2020, pp. 242–251.
[27] B. C. Benato, I. E. de Souza, F. L. Galvão, , and A. X. Falcão, "Convolutional neural networks from image markers," in *Beyond back-propagation: novel ideas for training neural architectures, Workshop at NeurIPS*, 2020.
[28] I. E. de Souza, B. C. Benato, and A. X. Falcão, "33rd conference on graphics, patterns and images. sibgrapi," in *Feature learning from image markers for object delineation*, 2020.
[29] L. M. João, M. C. Abrantes, B. C. Benato, and A. X. Falcão, "19th international joint conference on computer vision, imaging and computer graphics theory and applications. visapp (to appear)," in *Understanding marker-based normalization for FLIM Networks*, 2024.