Journal of Technological Information, Management & Engineering Sciences (2024),5(2), 82-94

Explainable Deep Clustering for Unsupervised Pattern Discovery in Complex Datasets

[1] **Dr. Waqas Ahmad**
[2] **Laiba Noor**

How to cite this article:

Ahmad, W., & Noor, L. (2024). Explainable Deep Clustering for Unsupervised Pattern Discovery in Complex Datasets. *Journal of Technological Information, Management & Engineering Sciences, 5(2), 82-94.*

Received	Accepted	Published
20 August 2024	22 October 2024	18 December 2024

Abstract

When there is no linear structure in the data, it is hard to discover patterns without supervision; if the data is high dimensional, discovery becomes more challenging; if there are heterogeneous variables in the data, the pattern discovery is more difficult; if there are ambiguous boundaries between groups, the discovery is more difficult. Deep clustering is a combination of representation learning and clustering to learn latent clusters without any predefined clusters, explainability is to make the learned patterns explainable and auditable. An integrative review of explainable deep clustering for complex data is presented, and the connection between representation learning, cluster assignment, dimensionality reduction, cluster validation, and explainability post hoc is highlighted. The review reveals a main gap – the patterns extracted with deep clustering techniques are frequently evaluated based on internal indices or external labels, while the semantic validity, stability, and human usefulness of the extracted patterns is less systematically addressed. The paper suggests a modular approach to use data auditing, representation learning, constrained clustering, structural exploration using UMAP, silhouette-based validation, feature-level interpretation, and stability analysis. Results are qualitative analytical results obtained from literature, not experimental results measured. The analysis shows that deep clustering can uncover a group structure that is nonlinear and cannot be captured by conventional distance-based clustering, and that embeddings and visualizations are not evidence of natural categories, in and of themselves. The proposed workflow will focus on reporting uncertainty, careful interpretation of cluster labels, domain validation, and reproducibility. Fuzzy clustering has been a well-known method for pattern discovery since it was first proposed.Fuzzy clustering is a well-known technique for pattern discovery since proposed that was popular before the advent of deep clustering techniques.

Keywords: *Deep Clustering, Explainable Artificial Intelligence, Representation Learning, Unsupervised Learning, Dimensionality Reduction, Cluster Validation, Fuzzy Clustering, UMAP Visualization, Cluster Stability, Pattern Discovery.*

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Introduction

Complex datasets often contain patterns that are not specified in advance. Behavioral segments of the customer population, patient sub-groups of medical measurements, operating regimes of sensor streams, and latent themes of text collections are examples. In order to overcome these issues, unsupervised clustering takes place which classifies observations based on their similarity, density, connectivity or learned representation. Clustering is an unsupervised classification technique, in which no labels are required, but rather structure is sought which can aid in later interpretation and decision-making.

The conventional clustering methods, such as k-means, are used to partition the observations based on an objective function which is based on distance. k-means, developed by MacQueen (1967), is a technique that divides the population into a number of given classes according to the criterion of minimizing the variance within each class. These types of methods are still useful for those groups with simple shapes and densities and simple manifold structure, but they may fail for groups with more complicated shapes and densities or with more complicated manifold structure.

These limitations are overcome by learning a representation in which it is easier to cluster, known as deep clustering. Xie et al. (2016) presented an algorithm named Deep Embedded Clustering (DEC) that learns feature representations and cluster assignments at the same time by a deep neural network. This can reveal the nonlinear structure, but also poses an interpretability problem: a cluster might be clearly separated in the learned embedding but not easily understood in terms of the original variables.

Explainable deep clustering aims to render the found groups understandable. It does not only ask how many clusters exist or how well separated they are but also what properties belong to each one, what is the stability of the clustering and if it is meaningful to domain experts. This paper summarizes some of the methods that have been proposed and suggests a workflow for responsible pattern discovery.

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Questions

The review consists of three questions. First, what makes deep clustering methods special in their ability to find complex patterns when compared to traditional clustering methods? Second, what is the role of dimensionality reduction, cluster validation and interpretation at the feature level with regard to explainability? Third, how can the uncertainty of clusters, or potentially their visual appeal, be guarded against in order to avoid overinterpretation? The paper is not asserting that any one deep clustering algorithm is better than any other.

Literature Review

Conventional Clustering Foundations

Clustering is an unsupervised learning problem where the observations are grouped according to a given notion of similarity. MacQueen (1967) cited the process of partitioning an N-dimensional population into k sets by a procedure called k-means, which was intended to divide the population in the most efficient manner with regard to within-class variance. K-means is still popular today because it is simple to compute and can yield centroids that can describe each group succinctly.

The k-means algorithm does, however, rely on several assumptions: Centroids can be reasonably used to represent clusters, and Euclidean distance is meaningful in a selected feature space. The following assumptions may not hold true for categorical data, text or images, networks, or datasets with non-linear boundaries. The method also assumes that the number of clusters is known beforehand and results may be sensitive to initialisation.

There is thus a need for cluster validation. Silhouettes were introduced as a graphical tool to aid the understanding and evaluation of cluster analysis by Rousseeuw (1987). A silhouette is a measure of how close a cluster is to the observation it has been assigned to, and how far it is from the closest other cluster. The average silhouette width can be used for comparison with different candidate clusterings, but cannot be used as a hard parameter for determining the importance of the cluster in the real world.

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Deep Embedded Clustering

Deep clustering is a combination of neural representation learning and clustering objectives. Xie et al. (2016) have introduced DEC, which iteratively minimizes cluster assignments and simultaneously learns a mapping from the original data space to a lower dimensional feature space. It is not a typical clustering algorithm applied directly to the high dimensional inputs, but uses the method to learn representations to make cluster structure more apparent.

One of the great benefits of deep clustering is that it can model nonlinear transformations that can separate groups that are overlapping in the original feature space, as shown in DEC. This is particularly applicable in the context of image, text, sensor and other complex data. Concurrently, the learned representation can mask the relationship between an embedding coordinate and an original variable, making the resulting clusters more difficult to interpret.

A deep clustering result is not a self-validating result, however. The fact that the reconstruction error is low, the clustering objective value is high or the embedding is visually distinguishable does not imply that the clusters represent meaningful real world categories. Validation should account for stability, domain relevance and the impact of preprocessing and hyperparameter selection.

Literature Review: Explainability

Dimensionality Reduction and Visualization.

Dimensionality reduction can facilitate exploration by reducing the dimensionality of the data. McInnes et al. (2018) introduced a nonlinear manifold-learning method called Uniform Manifold Approximation and Projection (UMAP) that is derived from Riemannian geometry and algebraic topology. The idea behind UMAP is to represent the high-dimensional data in a low dimensional space while attempting to capture some of the topological structure.

According to Healy and McInnes (2024), UMAP is a nonlinear dimension-reduction technique that is often used for visualisation and in a lot of other applications, including clustering. When using an explainable deep-clustering workflow, UMAP can assist the analysts in checking if clusters are clustered within sensible regions

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

of a learned embedding. But a 2-D representation is not a full picture of the data. The distance, the shape of clusters and the apparent separation between clusters can vary with hyperparameters and sampling.

In that case, visualization should not be considered as an evidence of cluster validity but as exploratory evidence. A good UMAP plot can exhibit structure, but can also emphasize separation, hide overlap, or be a preprocessing artifact. A solid practice is to look at several projections, settings, and cluster resolutions prior to arriving at a conclusion.

Interpreting Discovered Clusters

The need for explainability in clustering is the need to establish an association between abstract clusters and meaningful features. One way is to compare the centroids, medoids or summary statistics of the clusters in terms of the original variables. Another method is to train a transparent surrogate classifier that categorises members of one cluster from members of other clusters, and also analyse the importance of features for the surrogate. A third approach looks at representative examples, "border" cases and cluster-specific association rules. These methods can help to clarify clusters, but cannot eliminate the ambiguity. A feature can be used to define clusters statistically but not provide any explanations for how/why they exist. There may be several variables in correlated data sets that seem to be significant because they are measuring related underlying phenomena. Explanation should therefore be differentiated from description in interpretation.

Stability and Uncertainty

Cluster stability is an evaluation of how “stable” the cluster groupings are when the perturbations (random seeds, bootstrap resampling, feature subsets, preprocessing variants, and alternative numbers of clusters) are applied. Stable clusters are better candidates for further investigation than clusters which vanish with minor changes to the workflow. The stability of a pattern is not a guarantee of its semantic validity, but it does alert that a pattern can be unstable due to arbitrary choices.

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Uncertainty also occurs when looking at individual observations. Some observations may be strongly affiliated with one cluster, but others may be on a boundary or may have a potential for being in several clusters. Reporting the uncertainty about the membership of observations in clusters instead of giving each observation a single definitive label can enhance the transparency and responsible use of the clustering results.

Research Gap

Conventional clustering, deep representation learning, dimensionality reduction and cluster validation, are well-founded in the literature reviewed. However, there is still a lack of tools that can integrate these elements into a comprehensible and understandable process of complex datasets. The focus of deep clustering research tends to be clustering performance or embedding quality, or, for explainability research, individual prediction rather than unsupervised discovery at a group level.

The second gap is a separation between visual separation and substantive meaning. A high UMAP projection quality or high internal validation does not guarantee a cluster to be a useful segment, a disease subtype, an operational regime, a behavioral pattern. A third gap is the issue of the reproducibility of the clustering outcomes, such as initialization, preprocessing, hyperparameterization, and random sampling that can affect the results, but are not always reported systematically.

In this paper, this is addressed by proposing an integrated workflow where explainability is considered as a requirement at the pipeline level. The workflow is linked with representation learning, clustering, validation, visualization, feature-level interpretation, stability analysis. It is not the intent of this workflow to assert that this process has been substantiated in this review.

Objectives

The ultimate goal is to explore the potential of explainable deep clustering for discovering patterns in complex data without supervision. Specific objectives:

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2024),5(2), 82-94

Discuss about conventional clustering, deep embedded clustering, dimensionality reduction and cluster validation techniques.

- Show interpretability issues arising from deep clustering representations.

Outline an unsupervised pattern discovery process with a modular workflow that is transparent and reproducible.

- Outline the criteria of evaluation: structural quality, stability, semantic and human usefulness.
- Differentiate between results from literature and results that need to be tested with new experiments.

Working Proposition

The working hypothesis is that the quality of the representation, the structure of the learned clusters, their stability, and the interpretability of the resulting clusters in terms of the domain are all independent and must be evaluated together for unsupervised pattern discovery to be credible. The reviewed literature backs this proposition but it has not yet been tested experimentally in this paper. It needs to be assessed using controlled studies on a variety of datasets in the future.

Methodology

Review Design

This study applies an integrative review design with a set of selected foundational and influential scholarly sources which are purposively selected. MacQueen (1967) discusses k-means, Rousseeuw (1987) discusses silhouette validation, Xie et al. (2016) discuss Deep Embedded Clustering, McInnes et al. (2018) discuss UMAP, and Healy and McInnes (2024) discuss the use and interpretation of UMAP. The following sources are chosen because they are directly related to clustering, representation learning, validation or explainability.

The sources were analyzed for task, algorithmic mechanism, assumption, interpretations, and limitations. The review does not attempt to be comprehensive nor does it provide a systematic review, with database searches,

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

screening counts, or meta-analysis. Results obtained from the original papers are quantitative and not pooled since it is not possible to combine the data sets, evaluation protocols and computational settings.

Proposed Clustering Workflow

The proposed process has six steps: Problem formulation, Data audit, Data pre-processing, Representation learning, Clustering and Explanation. The first stage is to establish the discovery objective, the purpose of the use of clusters, and the criterion for deciding if the pattern is useful. In the second stage, missingness, duplication, outlier detection, leakage risks, class imbalance (when known), and group based dependencies are checked for data quality.

The goal of pre-processing is to convert the raw data into a form that is suitable to represent it while retaining its traceability. Only data that can be used for representation learning is used for scaling, imputation, encoding, normalization, and feature construction. The workflow then trains a lower dimensional representation with an autoencoder or other appropriate neural encoder. One existing method is Deep Embedded Clustering, where representation learning and cluster assignment are optimized concurrently (Xie et al., 2016).

The clustering stage makes multiple comparisons of algorithms and candidate numbers of clusters. A baseline using the k-means algorithm is kept for comparison, as it gives a straightforward point of comparison and generates centroid summaries. This is not presumed to be the best – it is evaluated as deep clustering. The assignments, membership ratings and boundary observations are focused on clusters instead of being decreased to one hard label.

Breadth and depth: Methodology: Validation and Explanation.

Structural Validation

Within the structural validation, internal measures are based on the concept of cohesion and separation. Silhouette analysis is suggested as a first validation tool since it measures the degree to which observations are closer to their assigned cluster than to adjacent clusters (Rousseeuw, 1987). The workflow also provides

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

average silhouette values and cluster level distributions to ensure that a high AVS does not mask for poorly resolved or ambiguous clusters.

The workflow also looks at the balance of cluster size, boundary density and sensitivity to cluster count. Don't choose the clustering solution just because it has the highest index in one of the internal clusters. Rather, the candidate solutions are contrasted through a variety of indices, visualization parameters and domain constraints. Results are reported under uncertainty and not in the form of definitive partitions.

Explainability Protocol

The result of the workflow for each cluster is a structured explanation, including representative examples, centroid or medoid summary, distinguishing variables and boundary cases. The use of UMAP as an exploratory visualization layer is used to observe if the learned representation shows coherent regions, but not as a replacement for validation in the original feature space (McInnes et al., 2018; Healy & McInnes, 2024).

Feature-level interpretation is used to compare cluster summaries and find variables that can be used to distinguish clusters by relying on a clear, transparent, and interpretable surrogate classifier. The variables are clustered for interpretation to prevent the assigning of importance to variables that are identical measures and not independent causes. Explanations are discussed with a rubric according to a pre-defined format with the help of domain experts and disagreements are not solved informally, but documented.

Stability and Reproducibility

The stability of the workflow is evaluated by rerunning the workflow with random seeds, bootstraps, various preprocessing options, and different counts of potential clusters. The more these perturbations are repeated the more credible a cluster will be and the more consistent the relationship with the original variables the more credible a cluster will be. The workflow also reports observations with changing assignments, as these cases may be cases of transition or ambivalence. Reproducibility needs versioning of data, preprocessing scripts, model configuration, random seeds, versions of software and explanations settings. All

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

the hyper parameters of embedding dimension, number of clusters, learning rate and initialization strategy are recorded. Any results should be reported with the exact configuration used to achieve that result.

Results

Qualitative Analytical Findings

The first analytical point is this: If we want to discover patterns in the unsupervised sense, we need to differentiate between what is group statistics and what is substantive meaning. The k-means offers a clear baseline based on the centroid (MacQueen, 1967), and silhouette analysis offers a structured approach to measuring cohesion and separation (Rousseeuw, 1987). A cluster in each case is not sufficient to suggest that it is a sensible real-world category.

The second is that deep clustering can be used to learn representations that facilitate clustering, and thus uncover nonlinear structure. Xie et al. (2016) illustrate this principle by using Deep Embedded Clustering, which simultaneously optimizes feature representation and cluster assignment. The interpretability cost is that the learned space may not be easily interpretable in terms of original variables.

The third finding is that visualization is a helpful way to explore, but not a way to validate. UMAP is meant to build low-dimensional representations that capture some of the topological structure (McInnes et al., 2018) and Healy and McInnes (2024) highlight its use for visualization and preprocessing. In a UMAP plot, any kind of separation must be viewed as an observation that suggests hypotheses, rather than as definitive proof of the existence of separate clusters.

The fourth finding: Explanability needs to be considered at cluster level as well as at the observation level. The explanation is helpful if it identifies some representative cases, distinguishes variables, observes boundary points and the extent to which the pattern is stable to perturbation. This is because clustering doesn't give you just one prediction with one local explanation.

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Proposed Outputs, Not Measurements

The methodological result is a specification for a transparent clustering process, but not a quantifiable improvement in performance. Deliverables are an audit on the data, a reproducible pipeline for representational learning, several clustering solutions, silhouette and stability reports, UMAP visualizations, explanation records per cluster level and a domain-validation log.

The authors do not report any silhouette scores, cluster counts, embedding plots, runtime measurements, or results of a user study as no experiments were run. The literature synthesis helps to develop the design of the proposed workflow, but any claim that the proposed workflow is superior to the current workflow in terms of discovery quality would need to be empirically tested.

Discussion

Interpretation and Practical implications.

The review proposes that deep clustering should not be a universal source of ground truth, but rather be employed as a discovery tool. Deep representations could expose patterns that would be hard to observe with the traditional distance-based approaches, but these patterns must be verified for use in segmentation, diagnosis, resource allocation, or decision making.

Simple baselines should be kept throughout analysis. While the deep model is more flexible, a result from k-means or hierarchical clustering can serve as a useful reference point that is interpretable. When performing a deep clustering solution, the resulting clusters should be substantially different, if so, the analysts should consider if the difference is due to genuine nonlinear structure, or due to any of the effects of the preprocessing, or due to instability of the learned representation.

This study found that the image may be limited, and that future research is required.

This paper suffers from a number of limitations. As the first, it is a focused review instead of an exhaustive survey of deep clustering, explainable artificial intelligence, or manifold learning. Second, it does not run any new experiments so

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

it cannot specify the relative performance of various clustering workflows. Third, the proposed methodology has been yet to be implemented or validated with real data.

The proposed workflow should be tested with future research on image, text, sensor, clinical and tabular datasets. Studies should include comparisons with standard baselines (e.g., linear baselines), where the preprocessing and validation conditions are matched, and the structural quality, stability, and computational cost measured, as well as the human interpretability. Other aspects that should be explored in research include the impact of embedding choices, the cluster-count selection, and visualization parameters on the semantic interpretation of groups that are discovered.

Conclusion

Explainable deep clustering is a promising way of unsupervised pattern discovery in complex datasets. It can make latent structure more accessible to analysts and domain expert by applying representation learning, clustering, dimensionality reduction, validation and feature-level interpretation. But the worth of a found cluster relies more on its internal validation score or visual separation than just that.

This paper presents an integrative review and a methodological proposal in modules that are not experimental results. The main idea of its paper is that you need transparency from the beginning to the end to get to a credible pattern discovery. Empirical research is needed in the future to validate the proposed workflow with reproducible pipelines, and with domain-expert validation and stability analysis.

References

Healy, J., & McInnes, J. (2024). Uniform manifold approximation and projection. Nature Reviews Methods Primers, 4(1), Article 82. https://doi.org/10.1038/s43586-024-00363-x

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

MacQueen, J. (1967). A few ways to classify and analyze multivariate observations. This is published in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 281–297). The University of California Press.

McInnes, L., Healy, J., & Melville, J. (2018). Uniform manifold approximation and projection for dimension reduction (UMAP). arXiv. https://arxiv.org/abs/1802.03426

Rousseeuw, P. J. (1987). Silhouettes: A graphical tool for interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Xie, J., Girshick, R., & Farhadi, A. (2016). Here we look at unsupervised deep embedding for clustering analysis. In the 33rd International Conference on Machine Learning, (pp.). 478–487). PMLR.

[1] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2] Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.