Interpretable Knowledge Discovery Reinforced by Visual Methods

Boris Kovalerchuk
Dept. of Computer Science
Central Washington University
USA
borisk@cwu.edu

ABSTRACT

This tutorial covers the state-of-the-art research, development, and applications in the KDD area of interpretable knowledge discovery reinforced by visual methods to stimulate and facilitate future work. It serves the KDD mission and objectives of gaining insight from the data. The topic is interdisciplinary bridging of scientific research and applied communities in KDD, Visual Analytics, Information Visualization, and HCI. This is a novel and fast growing area with significant applications, and potential.

First, in KDD, these studies have grown under the name of visual data mining. The recent growth under the names of deep visualization, and visual knowledge discovery, is motivated considerably by deep learning success in accuracy of prediction and its failure in explanation of the produced models without special interpretation efforts. In the areas of Visual Analytics, Information Visualization, and HCI, the increasing trend toward machine learning tasks, including deep learning, is also apparent.

This tutorial reviews progress in these areas with a comparative analysis of what each area brings to the joint table. The comparison includes the approaches: (1) to visualize Machine Learning (ML) models produced by the analytical ML methods, (2) to discover ML models by visual means, (3) to explain deep and other ML models by visual means, (4) to discover visual ML models assisted by analytical ML algorithms, (5) to discover analytical ML models assisted by visual means. . The presenter will use multiple relevant publications including his books: "Visual and Spatial Analysis: Advances in Visual Data Mining, Reasoning, and Problem Solving" (Springer, 2005), and "Visual Knowledge Discovery and Machine Learning" (Springer, 2018). The target audience of this tutorial consists of KDD researchers, graduate students, and practitioners with the basic knowledge of machine learning.

KDD '19, August 4–8, 2019, Anchorage, AK, USA.

ACM ISBN 978-1-4503-6201-6/19/08.
https://doi.org/10.1145/3292500.3332278

CCS CONCEPTS

• Machine learning • Human-centered computing • Visualization • Visual analytics

KEYWORDS

Visual knowledge discovery, data mining, explainable models, interpretability.

ACM Reference format:

Boris Kovalerchuk. 2019. Interpretable Knowledge Discovery Reinforced by Visual Methods. In Proceedings of KDD '19: The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, August 4–8,2019, 2 pages. https://doi.org/10.1145/3292500.3332278

1. TUTORIAL OUTLINE

The tutorial is structured as follows. It starts with the motivation then covers the approaches with case studies and finishes with the conclusion and future work. Visual Knowledge Discovery (VKD) is a computational paradigm based on visual representation of high-dimensional data, ML models and algorithms, which produce these models. Several studies show that a wise use of VDM can efficiently solve certain problems that are hard for pure analytical ML algorithms. This tutorial covers Visual Knowledge Discovery as the intersection of ML, HCI and Visual Analytics. The VKD includes different lines of research. This tutorial presents VKD aspects that describe and discover learning models with the use of visual representations. The main part of the tutorial on approaches and case studies covers:

- Visualization of ML models, produced by the analytical ML algorithms (visualization methods used to demonstrate ML models such as t-SNE, to show the superposition of input data/images with heat maps of model layers, dataflow graph in deep learning models, differences in the internal representations of objects, adversarial cases and others).
- Discovering ML models by visual means (lossless/reversible and lossy visual methods for n-D data representation based on Parallel and Radial Coordinates, RadVis, Manifolds, t-SNE General Line Coordinates, Shifted Paired Coordinates, Collocated Paired Coordinates, and others).
- Explaining ML models including deep learning models by visual means (activation and weight visualization, heat-map-based methods, dependency analysis, monotonicity approach

– monotone Boolean functions and chains; decision tree visualization, and others).

- Discovering visual ML models assisted by analytical ML algorithms, such as propositional and first order rules, random forests, CNN, decision trees, optimization based on genetic and other algorithms.

The tutorial is based on multiple relevant publications including those listed below. Figures below from [Kovalerchuk et al., 2005-2018] illustrate a part of the tutorial.

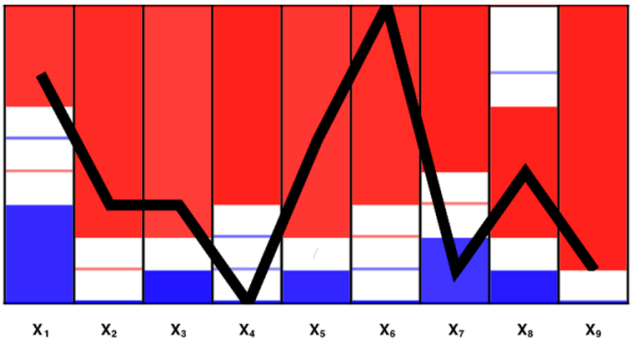

Figure 1: 9-D cancer case in parallel coordinates augmented with dominance intervals of classes.

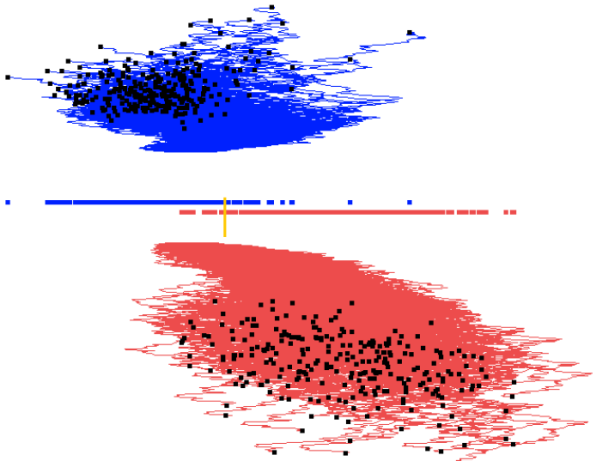

Figure 2: 484-D MNIST subset of images 22x22 for digits 0 and 1 in General Line Coordinates, GLC-L.

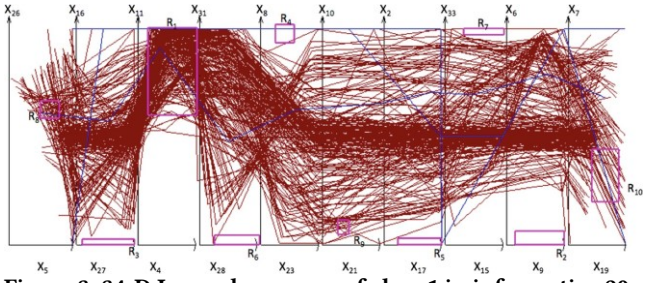

Figure 3: 34-D Ionosphere cases of class 1 in informative 20-D Shifted Paired Coordinates (SPC) subspace.

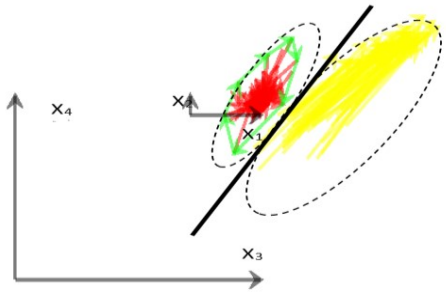

Figure 4: 4-D Iris data in SPC losslessly visualized and visually linearly discriminated.

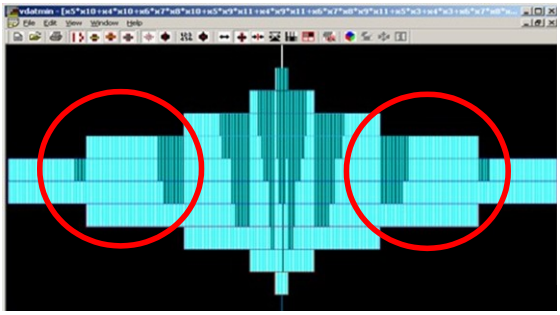

Figure 5: 11-D cancer cases in Monotone Boolean Space losslessly visualized.

REFERENCES

[1] D. Cashman, SR. Humayoun, F. Heimerl F, K. Park, S. Das, J. Thompson, B. Saket,, A. Mosca, J. Stasko, A. Endert, M.,Gleicher. Visual Analytics for Automated Model Discovery. arXiv preprint arXiv:1809.10782. 2018

[2] A.Endert, W.Ribarsky, C. Turkay, BW. Wong, I. Nabney, ID. Blanco, F. Rossi. The state of the art in integrating machine learning into visual analytics. In: Computer Graphics Forum 2017 Dec (Vol. 36, No. 8, pp. 458-486).

[3] Y. Guo, Y.Liu, A.Oerlemans, S. Lao, S. Wu, MS. Lew. Deep learning for visual understanding: A review. Neurocomputing. 2016 Apr 26;187:27-48.

[4] B. Kovalerchuk, V. Grishin. Adjustable General Line Coordinates for Visual Knowledge Discovery in n-D data, Information Visualization, 18(1),2019, 3-32.

[5] B. Kovalerchuk, Visual Knowledge Discovery and Machine Learning, Springer, 2018.

[6] B. Kovalerchuk, V. Grishin, Reversible Data Visualization to Support Machine Learning. In: S. Yamamoto and H. Mori (Eds.): Human Interface and the Management of Information. Interaction, Visualization, and Analytics, LNCS 10904, Springer, pp. 45-59, 2018.

[7] B. Kovalerchuk, A. Gharawi, Decreasing Occlusion and Increasing Explanation in Interactive Visual Knowledge Discovery, In: S. Yamamoto and H. Mori (Eds.) Human Interface and the Management of Information. Interaction, Visualization, and Analytics, LNCS 10904, Springer, pp. 505–526, 2018.

[8] B. Kovalerchuk, N. Neuhaus, Toward Efficient Automation of Interpretable Machine Learning. In: 2018 IEEE International Conference on Big Data, pp. 4933-4940, Seattle, Dec. 10-13, 2018 IEEE.

[9] B. Kovalerchuk, J. Schwing (eds), Visual and Spatial Analysis: Advances in Visual Data Mining, Reasoning, and Problem Solving, Springer, 2005,

[10] Danai Koutra, Di Jin, Yuanchi Ning, Christos Faloutsos. Perseus: An Interactive Large-Scale Graph Mining and Visualization Tool. Proc..of the VLDB Endowment, 2015. http://www.vldb.org/pvldb/vol8/p1924-koutra.pdf

[11] F. Poulet. Towards effective visual data mining with cooperative approaches. In: Visual Data Mining 2008 (pp. 389-406). Springer, Berlin, Heidelberg.

[12] D. Sacha, M. Kraus, DA. Keim, M. Chen. VIS4ML: An Ontology for Visual Analytics Assisted Machine Learning. IEEE transactions on visualization and computer graphics. 2019 Jan;25(1):385-95.

[13] C. Seifert, A. Aamir, A. Balagopalan, D. Jain, A. Sharma, S. Grottel, S. Gumhold, Visualizations of deep neural networks in computer vision: A survey. In: Transparent Data Mining for Big and Small Data 2017 (pp. 123-144). Springer,

[14] C. Turkay, R. Laramee, A. Holzinger, On the challenges and opportunities in visualization for machine learning and knowledge extraction: A research agenda. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction 2017 Aug 29 (pp. 191-198). Springer.