A Descriptive Framework for Stories of Algorithms

J. Liem[1], R. Henkin[1], J. Wood[1], and C. Turkay[1]

[1]City, University of London, United Kingdom

Abstract

Data-driven stories, widely used in journalism and scientific communication, match well with the recent focus on interpretable machine learning and AI explainability. Current technologies allow authors to break away from narratives that reflect traditional analytical workflows. To support designing such types of stories, we introduce a descriptive framework that helps identifying narrative patterns and other characteristics of algorithm-related stories. We describe the design space within the framework and demonstrate how to apply to an example of an algorithm-centered story, discussing potential future steps.

CCS Concepts

• ***Human-centered computing*** → ***Visualization theory, concepts and paradigms;*** *Visualization design and evaluation methods;*

1. Introduction and Background

Data-driven stories are now widely used in data journalism, business and scientific communication [NK15, RHDC18]. Besides helping to reveal *facts and insights* from data, storytelling techniques can also be used to explain the *methods* used during data analysis, such as detailing how an algorithm works or why a particular result was found. This type of communication is especially relevant for *interpretable machine learning* or *explainable AI*, which has received renewed interest from the visualization community [†]. Popular formats and platforms such as Jupyter Notebooks [KRKP*16] and R Markdown [AXM*19], and alternatives such as Observable [Obs17] and Idyll [CH18], allow users to create and share interactive documents intertwining narrative, code, and visualizations.

Some of these technologies allow authors to be more creative in their means of communication, breaking away from narratives that reflect traditional analytical workflows [BWF*18, RTH18]. Moreover, algorithms and models bring new levels of complexity regarding the abstractions and mechanisms employed. As such, new approaches that enable exploration of this complex space through creative storytelling techniques are required.

With this in mind, we propose a descriptive framework as a first step to help identify narrative patterns and other characteristics of algorithm-related stories. In contrast with existing taxonomies and design spaces, which focus on interaction aspects [MHL*17, SH10] or application of narrative theory to data-driven stories [RHDC18], our framework helps to analyze stories from an algorithm perspective, focusing on the way *the progression of an algorithm* is communicated and embedded in the narrative context.

[†] E.g. Workshop on Visualization for AI Explainability http://visxai.io

2. Descriptive framework

The framework's design space captures the visual expression of elements of models and algorithms in stories, relating this expression to the progress of an underlying computation. These computations include, for example, the application of an algorithm to some input data or the estimation of parameters of a model. The stories are assumed to contain the *explanation* of one or more elements of the algorithm in which a computation is involved, either live or pre-computed.

Each story is described across three axes: the visual expression of the different elements, capturing the algorithm perspective; the temporal aspects of these expressions, capturing the progression; and the interactive connections between the expressions, which represent the dynamic relationship between the visual expressions.

Axis 1 - Elements of algorithms and models:

Input I: these are the input records and features that are fed into an algorithm for building or applying a model.

Logic L: these are the internal structures and logical elements of algorithms, such as hidden layers in a neural network or centroids in clustering. Explanations of logic can also involve the use of different implementations in particular problems, such as showing how clustering algorithms arrive at a similar result.

Parameters P: these are the (hyper)parameters of algorithms and corresponding implementation, such as the number of clusters, the use of a regularization procedure or the number of hidden layers, among other choices.

Output O: parts of the story about the output aim to explain what models and algorithms *produce*, such as labels in a classification problem or a new set of dimensions in dimensionality reduction.

Axis 2 - Temporal aspects:

Progress depiction: an underlying computation can be depicted with *continuous* ( ) or *snapshot* ( ) representations. While a snapshot characterizes a single, static state of the algorithm, a continuous representation allows a viewer to follow the algorithmic progression in a more detailed uninterrupted way. Depending on the supported interactions, the progress can be perceived differently: a user can speed up or slow down the visual expression. We note that the progress of the depiction is not necessarily synchronized with the *computing time*, although this could also happen.

Thematic focus: in addition to the depiction of progress, we also describe the explanations of elements based on their relevance in each part of the story, which can be thematically in focus (/) or in a supporting role (/). The change in thematic focus drives the sequencing of parts of a story (1 2 3).

Axis 3 - Interactive connection

Element linkage: in interactive contexts, elements can also be can be linked ( ). In the framework, we map the interactions by connecting the sources and targets of transformations. The source is the focus of an interaction which is used to *indirectly* transform a target element. For example, interactions in the input indicate that any transformation of the input will affect the visual expression of the output. Depending on the degree of user involvement, such transformations can be triggered by user interactions (e.g. using a slider) or predefined by the author. In the visual description, we place the icon in the source of the transformation and color it by the target of the transformation.

3. Case study: A visual introduction to machine learning

A well known example is the *scrollstory* "A visual introduction to machine learning" [YC15], which explains the idea of machine learning using the methods of decision trees. The visual summary in Figure 1 demonstrates the application of the proposed framework.

Paced by the user, the machine learning model is explained and constructed step-by-step. The thematic focus moves from the input in part 1, which explains the records and features, to the logic in part 2, where the model logic is generated. Finally, the focus shifts to the output in part 3. Note that input and logic are still visually expressed through the moving dots and the tree structure, supporting the continuous representation of the model run as a whole. First, the training set is used, then the model runs again with a new test set. The change in the input, using different sets, is targeting the output, which is indicated in Figure 1 through the target symbol placed in the input row with the color of the output. In this example, no parameters, such as the *depth of a tree*, are addressed at all. In summary, the example follows a *build-up pattern*, which is suitable for explaining increasingly complex and interconnected elements along the logical order of algorithmic execution. Additional examples of stories and patterns are available in the supplementary materials and at https://algostories.github.io.

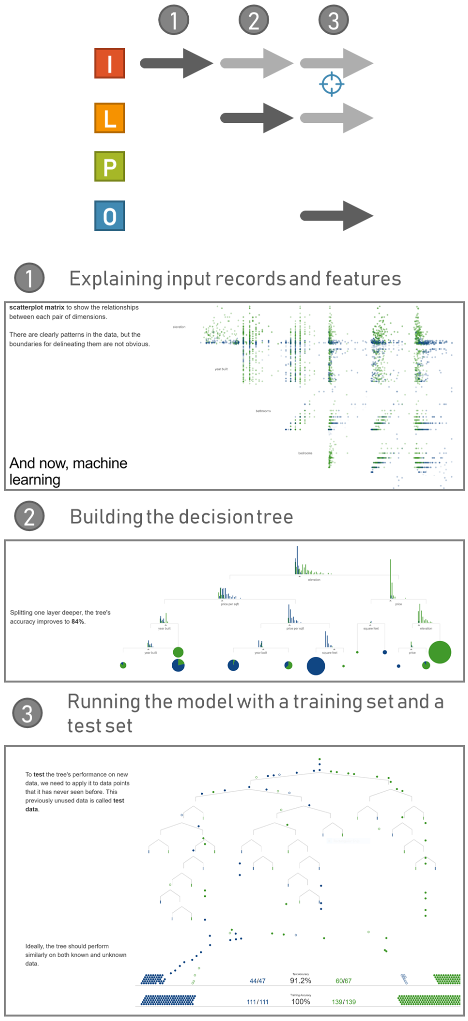

Figure 1: *Example "A visual introduction to machine learning" [YC15]*

4. Conclusion

We introduced a descriptive framework to help identify narrative patterns and other characteristics of algorithm-related stories and showed an example of how to apply it to existing stories. The identification of patterns enables a systematic critical review of story designs that can serve as *narrative templates*. The framework is an initial step towards a *generative* framework to help design algorithm-centred stories in a *platform-agnostic* manner. As future work, we also intend to expand from the limitation of linear or branching story structures to non-linear narratives and types of data where the progress of time plays an important role such as movement or streaming data.

Acknowledgements

JL and RH contributed equally to this work. RH and CT are supported by the UK Engineering and Physical Sciences Research Council (EPSRC) with grant number: EP/P025501/1.

References

[AXM*19] ALLAIRE J., XIE Y., MCPHERSON J., LURASCHI J., USHEY K., ATKINS A., WICKHAM H., CHENG J., CHANG W., IANNONE R.: *rmarkdown: Dynamic Documents for R*, 2019. R package version 1.12. URL: https://rmarkdown.rstudio.com. 1

[BWF*18] BACH B., WANG Z., FARINELLA M., MURRAY-RUST D., HENRY RICHE N.: Design patterns for data comics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, ACM, pp. 38:1–38:12. URL: http://doi.acm.org/10.1145/3173574.3173612, doi:10.1145/3173574.3173612. 1

[CH18] CONLEN M., HEER J.: Idyll: A markup language for authoring and publishing interactive articles on the web. In *ACM User Interface Software & Technology (UIST)* (2018). URL: http://idl.cs.washington.edu/papers/idyll. 1

[KRKP*16] KLUYVER T., RAGAN-KELLEY B., PÉREZ F., GRANGER B. E., BUSSONNIER M., FREDERIC J., KELLEY K., HAMRICK J. B., GROUT J., CORLAY S., IVANOV P., AVILA D., ABDALLA S., WILLING C., ET AL.: Jupyter notebooks - a publishing format for reproducible computational workflows. In *ELPUB* (2016). 1

[MHL*17] MCKENNA S., HENRY RICHE N., LEE B., BOY J., MEYER M.: Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. *Computer Graphics Forum 36*, 3 (2017), 377–387. doi:10.1111/cgf.13195. 1

[NK15] NUSSBAUMER KNAFLIC C.: *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley, Hoboken, New Jersey, 2015. 1

[Obs17] OBSERVABLE: Observable. https://observablehq.com/, 2017. Accessed: 2019-04-15. 1

[RHDC18] RICHE N. H., HURTER C., DIAKOPOULOS N., CARPENDALE S. (Eds.): *Data-Driven Storytelling*. CRC Press, Boca Raton, 2018. 1

[RTH18] RULE A., TABARD A., HOLLAN J. D.: Exploration and Explanation in Computational Notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (New York, New York, USA, 2018), ACM Press, pp. 1–12. URL: http://adamrule.com/files/papers/chi{_}2018{_}computational{_}notebooks{_}camera{_}ready.pdfhttp://dl.acm.org/citation.cfm?doid=3173574.3173606, doi:10.1145/3173574.3173606. 1

[SH10] SEGEL E., HEER J.: Narrative visualization: telling stories with data. *IEEE transactions on visualization and computer graphics 16*, 6 (2010), 1139–48. URL: https://ieeexplore.ieee.org/abstract/document/5613452/http://www.ncbi.nlm.nih.gov/pubmed/20975152, doi:10.1109/TVCG.2010.179. 1

[YC15] YEE S., CHU T.: A visual introduction to machine learning. http://www.r2d3.us/visual-intro-to-machine-learning-part-1/, 2015. Accessed: 2019-04-15. 2