JAvaScript Multimodal INformation Explorer

Fenja Schweder,^{1,2}, Sebastian Trujillo-Gomez², and Kai Polsterer,²

¹University of Bremen, Bremen, Germany; fen_sch@uni-bremen.de ²HITS gGmbH, Heidelberg, Germany

Abstract. Astronomical data is rich in volume, information and facets. Although this offers multiple research perspectives, processing the data remains a challenge. Infrastructures for analyzing, inspecting, exploring and communicating with data are mandatory. To address this issue, we introduce Jasmine, the JAvaScript Multimodal INformation Explorer. Jasmine allows users to open different data viewer modals that show a specific data point from a set. The viewer currently supports image data, as well as point cloud objects. Users can decide on which information about the data point they like to have displayed. Point clouds are interactive and allow for zooming, tossing, and turning. Picking a data point is enabled by providing a structured view of the set, arranged by a key property. This arrangement is achieved by autoencoding.

1. Introduction

Large surveys and extensive simulations sustain astronomers with information that is not only large in volume but also covers various aspects, fields, and dimensions. The effort in analyzing and processing the data is a remaining issue extensively described by Zhang & Zhao (2015) and is still under review (Faaique (2023)). This not only raises the expectation in algorithms to process and analyze the data but also requires infrastructures to apply them and work with the results.

Visualization and interactivity play a key role for the mentioned infrastructures. They have been addressed in earlier proposals for interactive exploration frameworks. Sciacca et al. (2014) introduced a platform to facilitate access, visualization, and exploration of big astronomical data. Another example is *Astronomaly* by Lochner & Bassett (2021), a web application detecting outliers in big data sets. *UltraPINK* (Kollasch & Polsterer 2024) visualizes common structures in a data set as Self-Organizing Kohonen Maps.

The mentioned solutions address the difficulty of providing a structured overview among a large data set. However, an additional challenge is displaying individual data points with multiple fields. A single visualization of a multidimensional data point is not capable of displaying all contained information without being overloaded. In addition, some aspects require different visualization techniques from others, due to their dimensionality or behavior. Desirable is a data inspector that offers the ability to switch between dimensions of interest.

2. The Concept of multimodal data exploration

In this work, we introduce the *JAvaScript Multimodal INformation Explorer*. JASMINE is a prototypical web application that enables inspecting and interacting with multidimensional data points from a big data set. The key feature of JASMINE is the gathering and comparing of different views on an astronomic object with multiple modals. In the following, we explain the underlying concept and how to implement the concept to address the requirements of astronomical data.

2.1. Designing a big data browser

The book *The Craft of Information Visualization* by Bederson & Shneiderman (2003) describes useful guidelines on how to create a user-friendly and comprehensible scientific visualization. Included in this work, Plaisant et al. (2003) provide comprehensive strategies for designing information browsers. Our approach of a multidimensional data explorer follows their taxonomy of multiple-view browsers.

Based on Shneiderman's mantra *Overview first, zoom and filter, then details-ondemand* (Shneiderman 2003), the launch window shows an overview among the whole data set. Selecting a data point will open an individual modal for the point where further and more detailed inspection is possible. Communication between modals is unidirectional from the overview window to the detail windows. By that means, interaction with the overview window may change the view in the detail window, but not vice versa.

Plaisant et al. (2003) recomment multi-window browsers with window placement and alignment strategies. Leaving the window management to the user also leaves them with additional workload. On the other hand, it is difficult to find an appropriate strategy that fulfills the user's needs for multiple cases. This is especially an issue with information browsers for explorative research, where the task focus may not be clear at first.

2.2. Massive and multivariate data

Since we expect the amount of data to be extremely large, an appropriate representation, if necessary aided by machine learning, is or can be chosen. Kollasch & Polsterer (2022) explain briefly how to apply techniques such as clustering, regression, and dimensionality reduction to create an interactive visualization of a big data set. The respective techniques are capable of creating an overview among the data, where strong features are outlined. The relations between the representation and the real data are preserved.

Showing astronomical objects with multiple fields in a single visualization is not possible without the visualization being overloaded. For this type of multivariate data, we designed a data type allowing one to store each field as individually accessible. Each data point from the original set is saved in this *data cube* format. Each side of the cube contains a different field that belongs to the underlying astronomical phenomenon. Data cubes can be nested to ensure a hierarchy among the properties, meaning that each side can have an arbitrarily large set of subsides. Users have the option to select a cube side to be shown in the detail window. The chosen side has to of the lowest hierarchy. Thanks to the multimodal approach, the cube sides and therefore the data fields can be directly compared.

3. Prototype

We sketch a prototypical implementation of the introduced concepts in JavaScript, together with a use-case example of a massive, multivariate data set. The result is a browser-driven web application that runs client-side. Against the recommendation of Plaisant et al. (2003), we did not implement a window management strategy, but instead use web browser windows as modals. Due to the variety, finding an appropriate strategy to automatically organize, align, and resize the windows requires specific attention and will be addressed in the future. Instead, the user is now free to design the browsing windows according to their requirements.

3.1. Example data and preparation

As a case of browsing extremely large multivariate data sets, we take snapshots from the *Illustris* project (Nelson et al. 2018) as an example. Illustris TNG is a set of large cosmological simulations that predict the formation and evolution of galaxies including many complex physical processes. The simulations model representative cosmological volumes across the history of the universe, following the evolution of multiple thousands galaxies. Each snapshot contains different types of particles. Each particle field has in turn multiple physical properties.

We took a subsample of 1000 galaxies that are provided with a stellar mock image. To generate an explorable surface, we use *Spherinator & HiPSter* framework (Polsterer et al. 2024). The *Spherinator* learns the intrinsic morphological features of galaxies from the mock images. It projects an arbitrarily large dataset using an interactive hierarchical spherical representation. The learned latent representation of galaxy morphological space is interpretable. The 3D positions in latent space reflect the morphological properties of each galaxy. To explore them, the *HiPSter* creates a hierarchical view as a HiPS tiling that can be examined with *Aladin Lite* (Baumann et al. 2022).

The last preparation step is organizing each galaxy from the sample as a stacked data cube with a side for each particle type and a subside for each data field, respectively. To do this on the fly with the HiPS tile generation, we apply light modifications to the *HiPSter* source code.

3.2. Dataset-driven methods

The interaction methods on the global dataset level are directly inherited from *Aladin Lite*. Zooming in increases the number of HealPiX cells and therefore the number of galaxies projected to the sphere surface. A dropdown menu allows one to change the view layer between the latent space reconstructions and the original images. Furthermore, it is possible to superimpose the catalog coordinates for all the objects and inspect the galaxy metadata interactively with access to the individual images and all the simulation data for that object directly through the TNG API.

Right clicking on a cell opens a detail window for the respective galaxy. The prototype can be configured to show either one detail window at a time or multiple windows of the same or different galaxies. The user can arrange, resize, and remove modals as required.

3.3. Datapoint-driven methods

The detail modal shows a selected subside of the data cube. By default, the mock image used to train the morphologic representations is shown. Since Illustris TNG provides

multiple particle clouds representing the gas, stellar, or dark-matter distribution, the user has the option to examine the 3D structures of the galaxy. JASMINE allows interacting with particle clouds via rotating and zooming. Color-coding shows the field values for the active particle clouds. A set of radio buttons lets the user determine which field should be shown. To make the difference clear, each field as a different color scale.

4. Conclusion and Outlook

JASMINE provides basic functionalities to explore a very large data set from a global to a detailed view point. The prototype shows the potential that comes with a multimodal data explorer and offers multiple ideas for future work. A window managing strategy suitable for mutable data would drastically increase the user experience. Bidirectional communication between detail windows and global view offers new possibilities to explore the data. Lastly, we plan for a generalized interface that allows us to attach JASMINE to any kind of data processing pipeline as an abstract data browser.

References

4

- Baumann, M., Boch, T., Pineau, F.-X., Fernique, P., Bot, C., & Allen, M. 2022, in Astronomical Data Analysis Software and Systems XXX, edited by J. E. Ruiz, F. Pierfedereci, & P. Teuben, vol. 532 of Astronomical Society of the Pacific Conference Series, 7
- Bederson, B. B., & Shneiderman, B. 2003, The Craft of Information Visualization (San Francisco: Morgan Kaufmann). URL https://www.sciencedirect.com/science/ article/pii/B9781558609150500573
- Faaique, M. 2023, International Journal of Mathematics, Statistics, and Computer Science, 2, 96–113. URL https://ijmscs.org/index.php/ijmscs/article/view/8561
- Kollasch, F., & Polsterer, K. 2022, in Astronomical Data Analysis Software and Systems XXX, edited by J. E. Ruiz, F. Pierfedereci, & P. Teuben, vol. 532 of Astronomical Society of the Pacific Conference Series, 11
- Kollasch, F., & Polsterer, K. 2024, Ultrapink new possibilities to explore self-organizing kohonen maps. 2406.03832, URL https://arxiv.org/abs/2406.03832
- Lochner, M., & Bassett, B. 2021, Astronomy and Computing, 36, 100481. URL http://dx. doi.org/10.1016/j.ascom.2021.100481
- Nelson, D., Pillepich, A., Springel, V., Weinberger, R., Hernquist, L., Pakmor, R., Genel, S., Torrey, P., Vogelsberger, M., Kauffmann, G., Marinacci, F., & Naiman, J. 2018, MN-RAS, 475, 624. 1707.03395
- Plaisant, C., Carr, D., & Shneiderman, B. 2003, in The Craft of Information Visualization, edited by B. B. Bederson, & B. Shneiderman (San Francisco: Morgan Kaufmann), Interactive Technologies, 353–363. URL https://www.sciencedirect.com/science/ article/pii/B9781558609150500457
- Polsterer, K. L., Doser, B., Fehlner, A., & Trujillo-Gomez, S. 2024. URL https://arxiv. org/abs/2406.03810
- Sciacca, E., Pistagna, C., Becciani, U., Costa, A., Massimino, P., Riggi, S., Vitello, F., Bandieramonte, M., & Krokos, M. 2014, in 2014 International Conference on High Performance Computing & Simulation (HPCS), 351
- Shneiderman, B. 2003, in The Craft of Information Visualization, edited by B. B. Bederson, & B. Shneiderman (San Francisco: Morgan Kaufmann), Interactive Technologies, 364–371. URL https://www.sciencedirect.com/science/article/pii/B9781558609150500469
- Zhang, Y., & Zhao, Y. 2015, Data Science Journal, 14