LEARNING STABLE REPRESENTATIONS IN A CHANGING WORLD WITH ON-LINE T-SNE: PROOF OF CONCEPT IN THE SONGBIRD

Stephane Deny
Department of Visual Information
Vision Institute
Paris, France
stephane.deny@inserm.fr

Emily Mackevicius, Tatsuo Okubo & Michale Fee
Department of Brain and Cognitive Sciences
McGovern Institute for Brain Research
MIT, USA
{elm,okubo,fee}@mit.edu

Gordon Berman
Department of Biology
Emory University, USA
gordon.berman@emory.edu

Joshua Shaevitz
Department of Physics
Princeton University, USA
shaevitz@princeton.edu

ABSTRACT

Many real-world time series involve repeated patterns that evolve gradually by following slow underlying trends. The evolution of relevant features prevents conventional learning methods from extracting representations that separate differing patterns while being consistent over the whole time series. Here, we present an unsupervised learning method to finding representations that are consistent over time and which separate patterns in non-stationary time-series. We developed an on-line version of t-Distributed Stochastic Neighbor Embedding (t-SNE). We apply t-SNE to the time series iteratively on a running window, and for each displacement of the window, we choose as the seed of the next embedding the final positions of the points obtained in the previous embedding. This process ensures consistency of the representation of slowly evolving patterns, while ensuring that the embedding at each step is optimally adapted to the current window. We apply this method to the song of the developing zebra finch, and we show that we are able to track multiple distinct syllables that are slowly emerging over multiple days, from babbling to the adult song stage.

1 INTRODUCTION

Conventional unsupervised learning algorithms rely on the underlying assumption that the relevant features to learn are stable across the whole dataset. However, in many real-world time series, the assumption of stationarity is compromised by slowly evolving trends (heteroscedasticity in financial time series, increasing popularity of a social website, emergence of syllables in bird song). Learning representations on the whole dataset risks reducing the separation between different patterns because of their respective evolution in time. Learning representations locally in time risks losing consistency of the representation across the different epochs.

Our method learns a slowly changing representation online by using a sliding window.

2 METHOD

Our method is based on t-distributed stochastic neighbor embedding (t-SNE), a recent technique for embedding high-dimensional data into a low-dimensional space that preserves relevant structure (Van der Maaten & Hinton (2008)).

We choose a time-window T, long enough so that each pattern of interest appears at least a few times in the window length, regardless of the position of the window in the time-series.

Then, we scroll the position of the window through the whole time series, with a time-step dt$<<$T. For each position of the window, we compute the t-SNE embedding of the data points in the time window. Note that each embedding is specific to the window on which it is computed, ensuring that it captures local structure optimally.

In order to compute t-SNE, we need to choose the initial coordinates of the data points in the low-dimensional space. Because dt$<<$T, two successive embeddings share most of their data points. For all the shared data points, we match their initial coordinates in the present embedding to their final coordinates in the previous embedding. This step ensures that the patterns are consistent across successive embeddings.

3 RESULTS

In order to prove the applicability of this method to real world non-stationary time-series, we applied it to samples of zebra finch song recorded over development.

Zebra finches sing a repertoire of distinct syllables that emerge during development from unstructured babbling, and have become a fruitful model for studying the neural mechanisms behind complex learned behaviors. Such studies require a precise classification of song itself. Previous semi-automatic methods for classifying song syllables involve clustering heuristic acoustic features (Tchernichovski et al. (2000)). However, clusters often overlap, especially in juvenile song, which has required manual classification of syllables from spectrograms (fig 1 A and B, Tchernichovski et al. (2001), Okubo et al. (2015)).

We applied our on-line t-SNE method to automatically extract song structure useful for classifying syllables and tracking their development.

Birdsong is produced in discrete syllables, separated by short gaps (fig 1A). As input to t-SNE, we used the song spectrogram (binned in 5ms slices, T = 10s, dt = 0.1s), as well as information about the location of each spectral slice within the syllable. Our embedding captured each syllable type as a distinct trajectory, well separated from that of other syllables (fig 1C). We made a GUI allowing users to label a small number of syllables, and these labels automatically propagate to the rest of the data (see video here: https://goo.gl/7XHkPk).

The embedding method allowed us to visualize the emergence of song structure during development (see video here: https://goo.gl/LgoJN8, dph = days post hatch, rotation added for clarity of visualization). We were able to visualize the progressive dissociation of a protosyllable into multiple distinct syllable types (fig1 D), consistent with previous observations of syllable differentiation (Tchernichovski et al. (2000), Okubo et al. (2015)).

4 DISCUSSION

Our on-line t-SNE method uses overlapping sliding windows to capture local structure, while maintaining consistency across time. We were able to represent song syllables as they emerged from babbling, then become adult songs. The learned representation was also useful for automatic syllable classification, a task which often previously involved tedious manual labeling. Our method could easily be adapted to efficiently and objectively classify events in other types of non-stationary time series.

Since our method is able to compactly represent the entity of song development, it opens up intriguing possibilities for quantifying cases of abnormal song development. Abnormal auditory experience, as well as certain brain manipulations, lead to pathological song (Scharff & Nottebohm

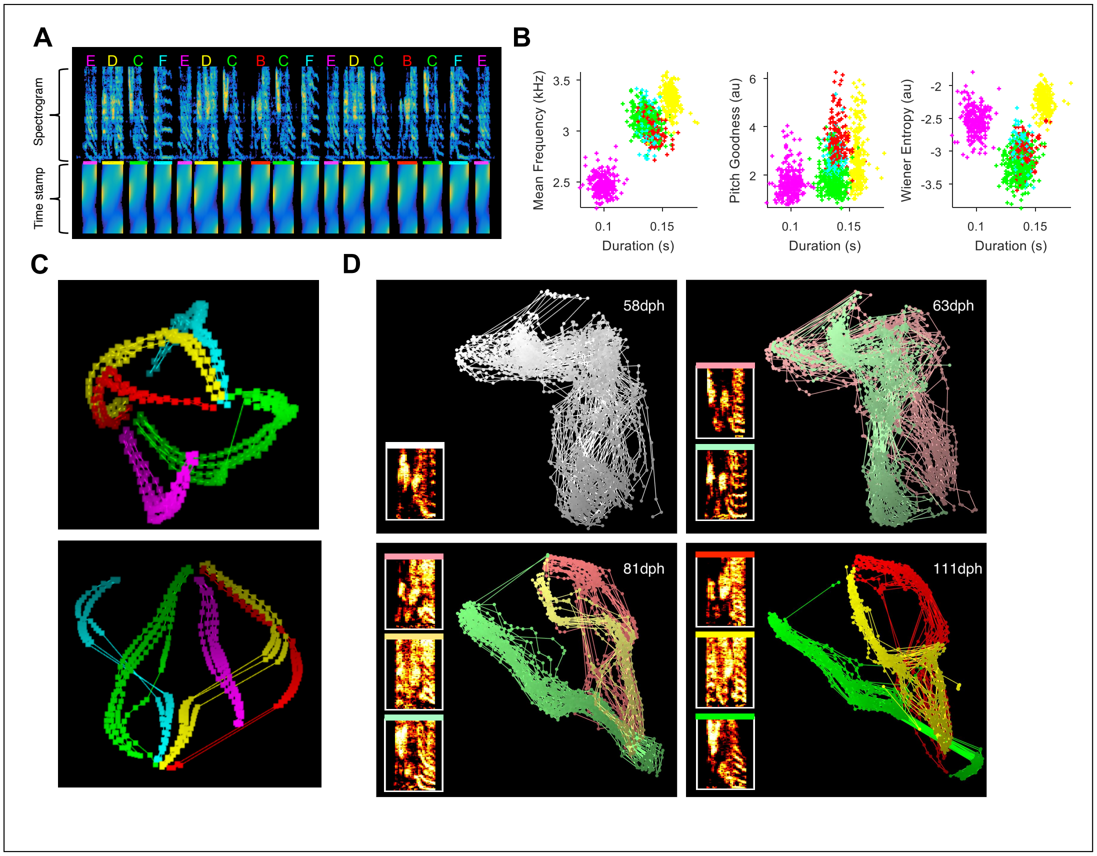

Figure 1: Extracting song structure with t-SNE: comparison with classical methods, and tracking emerging structure over development. We use a song where we have manual syllable labels as ground truth (Okubo et al. (2015)). A: Spectrogram of the song, for frequencies ranging from 0.5 to 6kHz, with manual syllable labels. Below the spectrogram is the time stamp code indicating the location of each spectral slice within the syllable (also used for t-SNE). B: Scatter plots of syllables, showing three different heuristic acoustic features versus syllable duration. Note that the cyan, green and red syllables are not well separated. C: (Top) t-SNE embedding of the song in a 2-dimensional space. Each dot represents a 5ms spectral slice. Each color corresponds to a different syllable (manual labeling), and saturation indicates the time of the slice within the syllable. Dots corresponding to consecutive slices are connected by a line. (Bottom) t-SNE embedding of the song in a 3-dimensional space. Note that t-SNE syllable trajectories form clusters consistent with manual labels. D: t-SNE embedding showing new syllables emerging during development from a common protosyllable. Trajectories at different ages are shown in separate panels, along with example spectrograms of each syllable type. Panels: (58 days post hatch, dph) protosyllable stage; (63 dph) emergence of two syllables, red and green indicate manual labeling; (81 dph) late juvenile stage showing further separation of the red and green syllables, and the emergence of a new syllable type, yellow; (111 dph) adult song.

(1991),Fehr et al. (2009)). Our new method provides a clear and compact representation of song development, which could help pinpoint how and when during development pathological songs first deviate from normal.

In the case of non-stationary time series, this on-line embedding method could be a useful alternative to a comprehensive embedding of the whole dataset. When a time series evolve to be very different from epochs to epochs, embedding all the data points in a single map will result in a less accurate picture of each epoch without adding relevant structure, since there is almost no useful structure to be kept across different epochs. From a computational point of view, it is a challenge to embed all the points of a huge dataset, because t-SNE has a O(n2) complexity. A recent study (Berman et al.

(2014)) has found a solution to embed huge datasets with t-SNE consisting in selecting a fraction of representative data points to shape the embedding space, but this selection still involves some computation overhead.

5 REFERENCES

REFERENCES

Gordon J. Berman, Daniel M. Choi, William Bialek, and Joshua W. Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99): 20140672, October 2014. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2014.0672. URL http://rsif.royalsocietypublishing.org/content/11/99/20140672.

Olga Fehr, Haibin Wang, Sigal Saar, Partha P. Mitra, and Ofer Tchernichovski. De novo establishment of wild-type song culture in the zebra finch. *Nature*, 459(7246):564–568, May 2009. ISSN 0028-0836. doi: 10.1038/nature07994. URL http://www.nature.com/nature/journal/v459/n7246/abs/nature07994.html.

Tatsuo S. Okubo, Emily L. Mackevicius, Hannah L. Payne, Galen F. Lynch, and Michale S. Fee. Growth and splitting of neural sequences in songbird vocal development. *Nature*, 528(7582):352–357, 2015. URL http://www.nature.com/nature/journal/v528/n7582/abs/nature15741.html.

C. Scharff and F. Nottebohm. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of Neuroscience*, 11(9):2896–2913, January 1991. ISSN 0270-6474, 1529-2401. URL http://www.jneurosci.org/content/11/9/2896.

Ofer Tchernichovski, Fernando Nottebohm, Ching Elizabeth Ho, Bijan Pesaran, and Partha Pratim Mitra. A procedure for an automated measurement of song similarity. *Animal Behaviour*, 59 (6):1167–1176, 2000. URL http://www.sciencedirect.com/science/article/pii/S0003347299914161.

Ofer Tchernichovski, Partha P. Mitra, Thierry Lints, and Fernando Nottebohm. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science*, 291(5513):2564–2569, 2001. URL http://science.sciencemag.org/content/291/5513/2564.short.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008. URL http://siplab.tudelft.nl/sites/default/files/vandermaaten08a.pdf.