Exploring chaotic time series and phase spaces

de Carvalho Pagliosa, Lucas

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
We now summarize the main findings on the research work described in this thesis, as follows.

Firstly, we have explored the relationship between entropy and the independence of phase states (Chapter 5). For this, we relied on the Statistical Learning Theory (SLT) framework (Section 5.2) to show some phase spaces (embedding with different parameters $m$ and $\tau$) led to better learning models after applying the same regression function. Further, as SLT requires the input space to be formed by independent-and-identically-distributed (i.i.d.) data, we empirically show that phase states have satisfied such an assumption after testifying generalization. For this study, it was assumed data to come from a controlled environment. Furthermore, we studied the correlation of entropy with different types of embeddings, both numerically (Chapter 5) and visually (Chapter 8). Although we found that optimal embeddings do have a low entropy, the relation was not one-to-one. For example, we found that very different embeddings from the known optimal one can also have low entropy levels. This raises the question of what is, actually, a good definition of an optimal embedding. Intuitively, one would say that deterministic systems generally have well-designed structures in the phase space, as one state maps to a single other in the future. Conversely, stochastic processes tend to have phase states spread all over the phase space, such that such patterns are rarer to happen. However, as outlined above, we could not find a one-to-one correspondence between the distance from an optimal embedding and the entropy level. This means, also, that using entropy as a criterion to compute the parameters of an optimal embedding is a very difficult, if even possible, task.

As part of our studies concerning chaos theory and dynamical systems, we wanted to prove the effectiveness of phase space models. We have performed experimental analysis in the context of time series semi-supervised classification, where we have compared time series and phase space methods using self-learning methods (Chapter 6). As shown there, our method based on state spaces yields more accurate results than state-of-the-art methods. As such, this strengthens our belief that modeling dynamical systems via their (optimal or near-optimal) embeddings in phase space is a useful proposition.

We next revisited the context of the SLT, aiming to come up with a set of criteria and methodology for qualitatively assessing...
how well learning is ensured by algorithms that attempt to detect concept drift in data streams (Chapter 7). We adapted the SLT to account for this more challenging type of data (as compared to time series drawn from a single phenomenon) and next evaluated several state-of-the-art concept-drift-detection algorithms. Saliently, our evaluation showed that no algorithm (from the studied ones) fully complies with our criteria, thus can ensure learning. This result points out some limitations of existing concept drift detection methods (which should be overcome by future research), although it also provides a new theoretical framework to evaluate all such algorithms in a principled way.

On a more practical side, we examined the challenge of getting a visual insight into large high-dimensional datasets such as created by our time series. We proposed RadViz++, a visual exploration tool that overcomes several limitations of existing data visualization tools based on the radial metaphor (Chapter 8). We validated this tool on several real-world high-dimensional datasets. Next, we tried to visually explore (in another proposal) the set of phase spaces generated by various embeddings. Our exploration revealed several interesting (though, complex to interpret) patterns, and correlated positively with the difficulty of fully characterizing optimal embeddings by low entropy levels.

Finally, we turned back to the problem of estimating optimal embeddings, and used artificial neural networks to address this task (Chapter 9). The proposed approach is simple to implement, fast to execute, has only a few parameters, and yields values for the optimal embedding parameters $m$ and $\tau$ which are very close to ground-truth values stated by the literature. Given these results, and the simplicity of our approach, we argue that exploring more refined deep-learning architectures is a promising way forward phase-space reconstruction.

Several future work is possible based on our results. From a theoretical viewpoint, it is interesting (and valuable) to further explore the relationship between entropy and optimal embeddings. By defining what an “optimal” embedding is (which may be a problem-specific question to answer), we believe that stronger correlations with entropy can be found. This aspect can benefit from refining and extending our phase-space visualizations to e.g., show not only how much two phase spaces are different, but what precisely makes them different. Along the same lines, methods to visually summarize large and complex trajectories in phase spaces are of interest. Also, ways to visualize trajectories in phase spaces having more than three dimensions, e.g., by the suitable use of dimensionality reduction, are an interesting extension to consider. At the practical endpoint, using different models fed by state-space
features could improve the classification and prediction of time series, better than with the current methods employed in this thesis.