An Extension of Dynamic Time Warping for Images: Dynamic Image Warping

Krisztian Buza **Margit Antal**

BioIntelligence Group, Department of Mathematics-Informatics
Sapientia Hungarian University of Transylvania
Targu Mures, Romania

buza@biointelligence.hu, manyi@ms.sapientia.ro

Dynamic time warping (DTW) is probably the most successful distance metric for time series classification. It has been used as an essential component of various time series classifiers, including neural networks, see e.g. [1] [2] [3] and the references therein. The core idea behind DTW can be summarized as follows: based on dynamic programming, DTW aims to find an optimal alignment between time series so that it allows for shiftings and elongations when it compares two time series. This is essential because in case of data captured under real-world conditions, we can not expect the same pattern to be repeated *exactly* (see e.g., the ECG of a patient, movements during handwriting or dynamics of keystrokes). A similar issue arises when we compare images or parts of images.

In this talk, we present a generalisation of dynamic time warping for the comparison of images. A digital image of $n \times m$ pixels may be seen as an intensity matrix with n rows and m columns. We consider such a matrix a sequence of n row vectors (or as a sequence of m column vectors, respectively). In principle, two sequences can be compared with DTW, however, we have to take into account that, in our case, items of the two sequences are not real numbers, but vectors. Therefore, we propose to use DTW as inner distance within the main DTW, i.e., for the comparison of the aforementioned row (or column) vectors. Thus, we can summarize our approach as *"DTW with DTW as inner distance"* or *"double DTW"*. For simplicity, we call it dynamic image warping or DIW for short.

We made a reference implementation (Python and Cython source codes) of our approach publicly available in our github repository:

https://github.com/kr7/diw

We performed initial experiments in the context of recognition of handwritten digits. The results show that, compared with other distance metrics, such as Euclidean or Manhattan distances, DIW is especially beneficial in cases when few training instances are available.

References

[1] Hiroaki Sakoe, Seibi Chiba: "Dynamic programming algorithm optimization for spoken word recognition" *IEEE transactions on acoustics, speech, and signal processing* 26.1 (1978): 43-49.

[2] Krisztian Buza: "Time series classification and its applications" *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics* 2018.

[3] Krisztian Buza, Margit Antal: "Convolutional neural networks with dynamic convolution for time series classification" *International Conference on Computational Collective Intelligence*, Springer, Cham, 2021.