Extracting Optimal Performance from Dynamic Time Warping

Abdullah Mueen
University of New Mexico
mueen@cs.unm.edu

Eamonn Keogh
University of California, Riverside
eamonn@cs.ucr.edu

ABSTRACT

Dynamic Time Warping (DTW) is a distance measure that compares two time series after optimally aligning them. DTW is being used for decades in thousands of academic and industrial projects despite the very expensive computational complexity, $O(n^2)$. These applications include data mining, image processing, signal processing, robotics and computer graphics among many others. In spite of all this research effort, there are many myths and misunderstanding about DTW in the literature, for example "it is too slow to be useful" or "the warping window size does not matter much."

In this tutorial, we correct these misunderstandings and we summarize the research efforts in optimizing both the efficiency and effectiveness of both the basic DTW algorithm, and of the higher-level algorithms that exploit DTW such as similarity search, clustering and classification. We will discuss variants of DTW such as constrained DTW, multidimensional DTW and asynchronous DTW, and optimization techniques such as lower bounding, early abandoning, run-length encoding, bounded approximation and hardware optimization.

We will discuss a multitude of application areas including physiological monitoring, social media mining, activity recognition and animal sound processing. The optimization techniques are generalizable to other domains on various data types and problems.

1. TOPIC OVERVIEW

With the advent of novel sensing techniques, time series data are ubiquitous. Large volumes of time series data are routinely created in scientific, industrial, entertainment, medical and biological domains. Examples include physiological recordings, space telemetry, performance counters, accelerometer signals, oceanographic signals, acoustics signals, quantified self, financial indices, etc.

Dynamic Time Warping (DTW) is a distance measure for time series data that is invariant to distortions in time. DTW has being used for decades in many application areas including speech processing, physiological monitoring, human activity recognition and financial engineering. Although time warping applications are ubiquitous, the computational complexity of DTW $O(n^2)$ is a major barrier to adoption. At least two hundred research papers has been published that focus on speeding up DTW algorithm and its applications to higher-level algorithms. The current fastest DTW algorithm can search a trillion subsequences in a day, or monitor streams in real-time [13], yet we argue that the full power of DTW remains underexploited.

In this tutorial, we focus on two key issues:

- How to do DTW efficiently
- How to do DTW effectively

Efficient: We will discuss many optimization techniques that have accelerated DTW algorithm and its usefulness in similarity search (in archived [6] and streaming data [16][13]), classification [3] and clustering (exact [18] and approximate [22]. The optimization techniques include early abandoning [13], lower bounding [8][9], bounded approximation [22], compressed comparison and hardware acceleration [17] among many others. These techniques have been successful used in many domains and are generalizable to many data types and problems commonly appearing in other domains. Thus attendees working on different data types (strings, graphs, images etc) will still come away with useful tools and ideas.

Effective: Many published applications of DTW do not achieve the quality of results that are possible, because the users make suboptimal choices. For example, not normalizing the data properly, not using the correct "warping window" parameter, not generalizing DTW to multidimensional data correctly, enforcing the "endpoint" constraint where it should not be used etc. We will explain these issues, and show how to squeeze the optimal performance out of DTW.

The tutorial will be illustrated with numerous real world examples created just for this tutorial, including case studies in physiology [20], social media, activity recognition [2], speech processing [11] and image processing [14]. We will end the tutorial with a set of open problems that are challenging and have potential to significantly advance the knowledge front. We will also discuss some closed problems that are either solved or do not exist anymore. The key technical challenges and available sources of data will be discussed.

KDD '16 August 13-17, 2016, San Francisco, CA, USA

ACM ISBN 978-1-4503-4232-2/16/08.

DOI: http://dx.doi.org/10.1145/2939672.2945383

2. COVERAGE AND RELATED WORK

The tutorial covers selected research published in top conferences (e.g. VLDB [6][8][19], SIGKDD [3][13][1][7], ICDE [9][16][21][2], ICDM [12][17][10] and SDM [4][18][22]). We will present a comparative picture of all these methods from different perspectives including efficiency, accuracy, interpretability, resource requirement and so on. In addition, there is considerable research work on time series that appears outside the mainstream data mining conferences; for example, much interesting work on time series is done in PERvasive technologies [5][20], computer vision (CVPR) [14] and speech processing (Tran. Acoustics) [11][15]. We have also made extensive efforts to summarize such work in our comparative study.

3. PRESENTER BIOS

Dr. Mueen is an Assistant Professor at Computer Science in University of New Mexico. He has won the runner-up of SIGKDD doctoral dissertation contest in 2012 and a best paper award in the same conference. His research work has been published in top conferences including SIGMOD, KDD and ICDM. He has recently given well received tutorials in SDM and ICDM conferences on repeated pattern mining from time series data.

Dr. Keogh is a prolific author in data mining, for example he is a top-ten most prolific author in all three of the top ranked data mining conferences, SIGKDD, ICDM and SDM. He has won best paper awards at ICDM, SIGKDD and SIGMOD. His H-index of 68 reflects the influence of time series representations and algorithms. He has given well-received tutorials at SIGKDD (three times), ICDM (four times), VLDB, SDM, and CIKM.

4. ACKNOWLEDGMENT

The material is based on work supported by the National Science Foundation Grant No. SHF:1527127.

5. REFERENCES

[1] D. J. Berndt and J. Clifford. Using Dynamic Time Warping to Find Patterns in Time Series. In *KDD Workshop*, pages 359–370, 1994.

[2] Y. Chen, G. Chen, K. Chen, and B. C. Ooi. Efficient processing of warping time series join of motion capture data. In *Proceedings - International Conference on Data Engineering*, pages 1048–1059, 2009.

[3] Y. Chen, B. Hu, E. Keogh, and G. E. Batista. DTW-D: Time Series Semi-supervised Learning from a Single Example. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*, page 383, New York, New York, USA, Aug. 2013. ACM Press.

[4] S. Chu, E. Keogh, D. Hart, and M. Pazzani. *Iterative Deepening Dynamic Time Warping for Time Series*, chapter 12, pages 195–212. 2002.

[5] P. Jangyodsuk, C. Conly, and V. Athitsos. Sign Language Recognition Using Dynamic Time Warping and Hand Shape Distance Based on Histogram of Oriented Gradient Features. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '14, pages 50:1—-50:6, New York, NY, USA, 2014. ACM.

[6] E. Keogh. Exact indexing of dynamic time warping. In *Proceedings of the 28th international conference on Very Large Data Bases*, VLDB '02, pages 406–417, 2002.

[7] E. J. Keogh and M. J. Pazzani. Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00*, pages 285–289, New York, New York, USA, Aug. 2000. ACM Press.

[8] E. J. Keogh, L. Wei, X. Xi, S.-h. Lee, and M. Vlachos. LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures. In *Vldb*, pages 882–893, 2006.

[9] S.-W. Kim, S. Park, and W. W. Chu. An index-based approach for similarity search supporting time warping in large sequence databases. In *Data Engineering, 2001. Proceedings. 17th International Conference on*, pages 607–614, 2001.

[10] A. Mueen. Enumeration of time series motifs of all lengths. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, ICDM, pages 547–556, 2013.

[11] C. Myers, L. Rabiner, and A. Rosenberg. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(6):623–635, Dec. 1980.

[12] F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh. Dynamic Time Warping Averaging of Time Series Allows Faster and More Accurate Classification. In *2014 IEEE International Conference on Data Mining*, pages 470–479. IEEE, Dec. 2014.

[13] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 262–270, 2012.

[14] T. M. Rath and R. Manmatha. Word image matching using dynamic time warping. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II—-521. IEEE, 2003.

[15] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, Feb. 1978.

[16] Y. Sakurai and C. Faloutsos. Stream Monitoring under the TimeWarping Distance. *Carnegie Mellon University Research Showcase*, 0:1046–1055, 2007.

[17] D. Sart, A. Mueen, W. Najjar, V. Niennattrakul, and E. Keogh. Accelerating Dynamic Time Warping Subsequnce Search with GPUs and FPGAs. ICDM 2010. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 1001–1006, 2010.

[18] M. Shokoohi-Yekta, J. Wang, and E. Keogh. *On the Non-Trivial Generalization of Dynamic Time Warping to the Multi-Dimensional Case*, chapter 33, pages 289–297.

[19] M. Toyoda, Y. Sakurai, and Y. Ishikawa. Pattern Discovery in Data Streams Under the Time Warping Distance. *The VLDB Journal*, 22(3):295–318, June 2013.

[20] N. Tselas and P. Papapetrou. Benchmarking Dynamic Time Warping on Nearest Neighbor Classification of Electrocardiograms. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '14, pages 4:1—-4:4, New York, NY, USA, 2014. ACM.

[21] B.-K. Y. B.-K. Yi, H. Jagadish, and C. Faloutsos. Efficient retrieval of similar time sequences under time warping. In *Proceedings 14th International Conference on Data Engineering*, pages 201–208, 1998.

[22] Q. Zhu, G. Batista, T. Rakthanmanon, and E. Keogh. A Novel Approximation to Dynamic Time Warping allows Anytime Clustering of Massive Time Series Datasets. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 999–1010, 2012.