Similarity Search of Time Series Trajectories Based on Shape

Gajanan Gawde
Goa College of Engineering
Farmagudi
Goa-India 403401
gg@gec.ac.in

Jyoti Pawar
Goa University
Taleigao Plateau
Goa-India 403206
jdp@unigoa.ac.in

ABSTRACT

Many researchers have been attracted towards searching of similar moving objects' trajectories due to their myriad applications. Similarity of trajectories is determined mainly by closeness or shape parameters. Closeness is measured by distance measures like Dynamic Time Warping (DTW), Edit Distance with Real Penalty (ERP), Edit Distance with Real Sequence (EDR) and Longest Common Subsequence (LCSS). These edit distance measures support scaling and translation property, but do not support rotation invariant property. Rotation Invariant (RI) distance measure supports all three properties and hence it is considered to be superior compared to other edit distance measures. However, it has two main drawbacks. The first one is, RI does not compares trajectories based on the shape. The second drawback is, RI is not robust to noise and hence produces poor results. In order to eliminate drawbacks faced by existing measures, in this paper, we have proposed Polygon Based Distance (PBD) measure to compare trajectories for similarity. Our proposed PBD distance measure supports scaling and translation properties like other measures. The main advantages of PBD distance measure are, it is invariant of rotation, it compares the trajectories based on the shape and it is robust to noise. We have performed experimental study on real time and synthetic datasets. The average accuracy of DTW is 21.67, RI is 35 and PBD is 58.33. Since PBD measure compares the trajectories based on shape and is robust to noise, its accuracy is higher compared to DTW and RI distance measures.

KEYWORDS

Time series Trajectories, Similarity Search of Trajectories, Polygon Based Distance Measure, Shape Based Similarity Search

ACM Reference Format:
Gajanan Gawde and Jyoti Pawar. 2018. Similarity Search of Time Series Trajectories Based on Shape. In *CoDS-COMAD '18: The ACM India Joint International Conference on Data Science & Management of Data, January 11–13, 2018, Goa, India.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3152494.3167986

1 INTRODUCTION

The path traversed by a moving object is called as a trajectory. Trajectory is represented in two dimensional space using latitude and longitude. The general form of two dimensional trajectory representation is (oid,t,x,y) where oid is object identification, t is the time at which coordinates were recorded, x is the latitude and y is the longitude. There are a wide range of real time applications involving searching of similar trajectories. Some of them are Route Planning, Popular Route Identification, Car Pooling, Stock Exchange Retrieval, ECG Retrieval, Personal Security etc.

Similarity of trajectories is determined mainly by closeness or shape parameters. Closeness is measured by using a variety of distance measures. Dynamic Time Warping (DTW), Edit Distance with Real Penalty (ERP), Edit Distance with Real Sequence (EDR) and Longest Common Subsequence (LCSS) are popular distance measures based on edit distance. These existing edit distance measures are invariant of scaling and translation but not invariant of rotation. Rotation Invariant (RI) distance measure is invariant of all three properties and as such is considered to be the best. However, Rotation invariant distance measure does not support shape based comparison of trajectories and it is very sensitive towards noise. We have provided a motivational example to discuss the drawbacks faced by RI distance measure.

Example :- Consider the trajectories as shown in Figure 1. Query trajectory *Q* is compared with three trajectories for similarity using RI and DTW measures. The result of comparison is shown in Table 1. Query trajectory *Q* is similar to *T1*, *T2* and *T3* trajectories based on shape. However, similarity distance of RI and DTW are quite high and do not show any kind of similarity between the trajectories. This is because RI and DTW distance measures do not compare trajectories based on shape.

So, there is a need to compare trajectories based on shape. In this paper we have proposed Polygon Based Distance measure in order to tackle the problems faced by DTW and RI distance measures. Polygon Based Distance measure compares trajectories based on shape of the trajectories and it is robust to noise. Our experimental results show that, PBD measure performs better since it compares two trajectories based on shape and it is robust to noise.

2 RELATED WORK

Searching of similar time series trajectories is quite an interesting problem and quite a few researchers have contributed towards this field. The pioneering work by [1], [2] compares trajectories for similarity using Euclidean distance measure. The time series are scaled and transformed efficiently and compared for similarity in [10]. [4] introduced DTW to allow time series trajectories to be stretched to provide better match. The performance of DTW was highlighted in [18] and subsequently enhanced using lower bounding distance measure with segmentation. [9] introduced bounded similarity query with the help of Euclidean distance to identify similar time series trajectories. The time series trajectories are shifted, scaled

CoDS-COMAD '18, January 11–13, 2018, Goa, India

ACM ISBN 978-1-4503-6341-9/18/01…$15.00
https://doi.org/10.1145/3152494.3167986

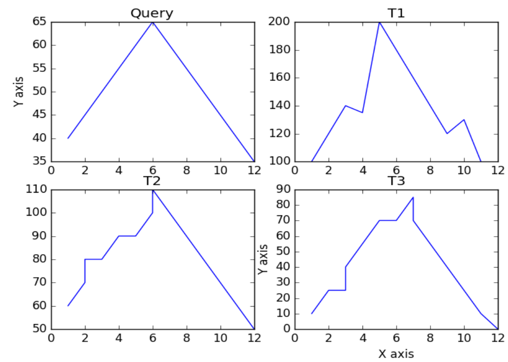

Figure 1: Stock Exchange Trajectories - Dataset 1

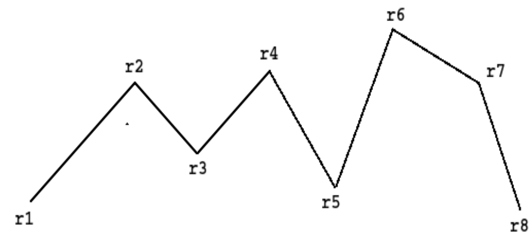

Figure 2: Trajectory Partition

	(Q,T1)	(Q,T2)	(Q,T3)
DTW	35.58	58.28	46.07
RI	257	222	247

Table 1: Similarity Distance - Dataset 1

and then similarity distance is computed. This is then compared with a threshold value, to identify similar time series.

In [8] and [17], LCSS measure is used to compare trajectories for similarity. LCSS measure allows a variable length gap to be inserted during matching of the trajectories and hence is robust to noise. In [5] modified LCSS measure is defined to compare trajectories of the objects from video. First 3D trajectories are extracted from a video and then LCSS distance measure is used to compare trajectories. Searching of similar multi dimensional trajectories was explored in [17]. Authors have used LCSS distance measure to compare multi dimensional trajectories. LCSS distance measure is used in conjunction with sigmoidal function in [15] to compare trajectories for similarity. In [16], indexed multi dimensional trajectories supporting multiple distance measures such as LCSS, DTW are proposed and index structure was designed in a such a way that there was no need to rebuild index again and again. In [11] various dimensionality reduction methods were investigated. It also contributed novel Piecewise Aggregate Approximation (PAA) technique to reduce the dimensionality. [13] highlights the extra computation done by LCSS and same is enhanced by fine tuning a threshold value.

In [6], ERP distance measure is proposed to compare trajectories for similarity. In ERP measure gap penalty is introduced to improve the performance of distance measure. ERP distance measure supports metric property which helps in pruning the distance computations. In [7], EDR distance measure is proposed to compare trajectories for similarity. EDR distance measure is discretized to 0 and 1 based on matching of the points. In [14], Rotation Invariant distance measure is proposed to compare time series trajectories. RI distance measure is invariant of scaling, translation and rotation. In [3], polygon distance function is proposed to compare polygonal shapes. The polygon distance function compares polygons based on the turning functions of two polygons. It compares polygons based on their shape and is robust to noise.

3 POLYGON BASED DISTANCE(PBD) MEASURE FOR TIME SERIES TRAJECTORIES

The time series trajectories are preprocessed to identify valid polygons and then converted into polygonal representation. The processed trajectories are then compared for similarity using PBD distance measure. PBD distance measure is defined using polygonal turning function defined in [3].

3.1 Pre-processing of Trajectories to Extract Valid polygons

Let polygon list $L = \{p_1, p_2, p_3, \ldots, p_N\}$ consist of N number of points of a trajectory. The polygon of a trajectory is said to be a Valid polygon if and only if segment joining p_1 and p_N points of the polygon list is not intersecting or overlapping with any other segments of the polygon and should have maximum number of segments.

Steps to Extract Valid polygons from Trajectory

(1) Let R be a trajectory with N number of points such as $\{r_1, r_2, r_3, \ldots, r_N\}$.
(2) Consider the first three points of trajectory R and add them to polygon list L and make it as first polygon i.e $L = \{r_1, r_2, r_3\}$. Check if L satisfies Valid polygon condition.
(3) If L satisfies Valid polygon condition, then add next point r_4 to the list L. Check if L is Valid polygon. If yes, then add next point to the list L and again check the condition. Repeat this process till L fails to satisfy condition. If L fails to satisfy Valid polygon condition after exploring all points from trajectories, then declare list L as Valid polygon and there is only one valid polygon for given trajectory. If L is not Valid polygon, then goto step 4.
(4) If L is not a Valid polygon, remove the last point appended to the list L. Now check the size of list L and if number of segments are equal or more than two, then declare L as the valid polygon. Reset the content of list L and add next three points to the list L which are yet to be explored. Goto step 3. If size of the list L is less than three, then remove first point of list L and add next two points to the list L. Goto step 3.

(5) Repeat the process till all the points of trajectory are explored.

The pre-processing algorithm is applied to the trajectory in Figure 2 and one Valid polygon $Pol_1 = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8\}$ is extracted.

3.2 Polygon Based Distance measure : PBD

Given two trajectories R and S with m and n number of polygons in it. Polygon Based Distance measure compares the two trajectories R and S for similarity by comparing polygons from both trajectories. Two trajectories are compared exhaustively with all possible combination of the polygons from the trajectories. Polygon Based Distance measure is defined as follows:

$$PBD(R,S) = \begin{cases} min\{(PBD(RestPoly(R), RestPoly(S))+ \\ MatchPoly, PBD(RestPoly(R), S) + 1, \\ PBD(R, RestPoly(S) + 1))\} & \text{if } m \geq 1 \text{ and } n \geq 1 \\ n, & \text{if } m = 0 \\ m, & \text{if } n = 0 \end{cases}$$

MatchPoly function compares two polygons using turning function. Two polygons are said to be similar if similarity distance of two polygons is less than or equal to $\delta 1$, where $\delta 1$ is a user threshold value. In paper [3], authors have proposed that best threshold value to match two polygons is 0.5. In the case of noisy datasets, the best threshold value is 0.8. Every time two polygons are compared, similarity distance is updated either by 1 or 0 based on whether the polygons are similar or not.

3.3 Problem Statement

Given Query trajectory $Q = \{(t_1, q_1), (t_2, q_2), \ldots, (t_M, q_M)\}$, Database trajectories $D = \{(d_1, d_2, \ldots, d_{NT})\}$ where $\{d_1, d_2, \ldots, d_{NT}\}$ are database trajectories and a user specified thresholds $\delta 1$ and $\delta 2$. The $\delta 1$ threshold value is used to compare polygons for similarity, where as $\delta 2$ threshold value is used for similarity distance of the trajectories and NT is number of database trajectories. The query trajectory is compared with database trajectories for similarity based on shape of the trajectories. The two trajectories Q and S are said to be similar if and only if similarity distance returned by PBD(Q,S) measure is less than or equal to $\delta 2$ where Q is query trajectory and S belongs to D.

4 EXPERIMENTAL STUDY

PBD distance measure is applied to Dataset 1 and results are as shown in Table 2. PBD distance measure is showing similarity match of Query trajectory Q with T1, T2 and T3 trajectories. This is because PBD distance measure compares trajectories based on shape. Accuracy of DTW, RI and PBD distance measures was recorded for different datasets as shown in Table 3. The average accuracy of DTW is 21.67, RI is 35 and PBD is 58.33 as shown in Table 3. PBD measure compares the trajectories based on shape and is robust to noise, thus its accuracy is at a higher side compared to DTW and RI distance measures. Stock Exchange Dataset1 is taken from [12] and Stock Exchange Dataset2 is from www.quandl.com.

	(Q,T1)	(Q,T2)	(Q,T3)
DTW	35.58	58.28	46.07
RI	257	222	247
PBD	1	0	1

Table 2: DTW, RI and PBD Distances on Dataset1

Datasets (Out of 30 correct samples)	DTW	RI	PBD
Stock Exchange Dataset1	7	10	18
Stock Exchange Dataset2	6	11	17

Table 3: Accuracy of Distance Measures

5 CONCLUSION

We have proposed Polygon Based Distance measure to compare time series trajectories for similarity. PBD distance measure compares trajectories based on shape of trajectories and is invariant of scaling, translation and rotation. PBD distance measure is compared with RI and DTW algorithms experimentally. Experimental study is carried out on real time and synthetic datasets of time series trajectories. Our experimental results show that the average accuracy of DTW is 21.67, RI is 35 and PBD is 58.33. Since PBD measure compares trajectories based on shape and is robust to noise, its accuracy is at a higher side compared to DTW and RI distance measures.

6 ACKNOWLEDGMENT

The authors are grateful to Mr. Chetan Desai for his critical review, encouragement and insightful discussions. The authors acknowledge the financial support received from Government of Goa vide Order No.16/ 231/ Hr.Studies/ GEC/ DTE/ 08/ Vol.I/ 467. They also thank the Department of Computer Science and Technology, Goa University and Department of Computer Engineering, Goa College of Engineering for unstinting support.

REFERENCES

[1] Rakesh Agrawal, Christos Faloutsos, and Arun Swami. 1993. Efficient Similarity Search In Sequence Databases. In *FODO*. Springer Verlag, 69–84.
[2] Rakesh Agrawal, King ip Lin, Harpreet S. Sawhney, and Kyuseok Shim. 1995. Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. In *In-VLDB*. 490–501.
[3] Esther M. Arkin, L. Paul Chew, Daniel P. Huttenlocher, Klara Kedem, and Joseph S. B. Mitchell. 1990. An Efficiently Computable Metric for Comparing Polygonal Shapes.. In *SODA*, David S. Johnson (Ed.). SIAM, 129–137.
[4] Donald J. Berndt and James Clifford. 1996. Finding Patterns in Time Series: A Dynamic Programming Approach. In *Advances in Knowledge Discovery and Data Mining*. 229–248.
[5] Dan Buzan, Stan Sclaroff, and George Kollios. 2004. Extraction and Clustering of Motion Trajectories in Video. In *17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, August 23-26, 2004*. 521–524.
[6] Lei Chen and Raymond T. Ng. 2004. On The Marriage of Lp-norms and Edit Distance.. In *VLDB* (2005-04-12), Mario A. Nascimento, M. Tamer ÃŰzsu, Donald Kossmann, RenÃľe J. Miller, JosÃľ A. Blakeley, and K. Bernhard Schiefer (Eds.). Morgan Kaufmann, 792–803.
[7] Lei Chen and M. Tamer ÃŰzsu. 2005. Robust and fast similarity search for moving object trajectories. In *In SIGMOD*. 491–502.

[8] Gautam Das, Dimitrios Gunopulos, and Heikki Mannila. 2001. Time-Series Similarity Problems and Well-Separated Geometric Sets. *Nord. J. Comput.* 8, 4 (2001), 409–423.

[9] Dina Q. Goldin, Todd D. Millstein, and Ayferi Kutlu. 2004. Bounded similarity querying for time-series data. *Information and Computation* 194, 2 (2004), 203 – 241. Special Issue Commemorating the 50th Birthday Anniversary of Paris C. Kanellakis.

[10] Man H. Wong Kelvin Kam Wing Chu. 1999. Fast Time Series Searching with Scaling and Shifting. In *ACM-PODS*. 237–248.

[11] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. 2000. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. *JOURNAL OF KNOWLEDGE AND INFORMATION SYSTEMS* 3 (2000), 263–286.

[12] M. Lichman. 2013. UCI Machine Learning Repository. (2013).

[13] Michael D. Morse and Jignesh M. Patel. 2007. An efficient and accurate method for evaluating time series similarity.. In *SIGMOD Conference* (2007-06-13), Chee Yong Chan, Beng Chin Ooi, and Aoying Zhou (Eds.). ACM, 569–580.

[14] Michail Vlachos. 2004. Rotation invariant distance measures for trajectories. In *Proceedings of SIGKDD*. 707–712.

[15] Michail Vlachos, Dimitrios Gunopulos, and George Kollios. 2002. Robust Similarity Measures for Mobile Object Trajectories.. In *DEXA Workshops* (2002-11-04). IEEE Computer Society, 721–728.

[16] Michail Vlachos, Marios Hadjieleftheriou, Dimitrios Gunopulos, and Eamonn J. Keogh. 2006. Indexing MultiÂŋDimensional TimeÂŋSeries with Support for Multiple Distance Measures. *VLDB J.* 15, 1 (2006), 1–20.

[17] Michail Vlachos, George Kollios, and Dimitrios Gunopulos. 2002. Discovering Similar Multidimensional Trajectories. In *In ICDE*. 673–684.

[18] Christos Faloutsos Yasushi Sakurai, Masatoshi Yoshikawa. 2005. Fast Similarity Search under the Time Warping Distance. In *ACM-PODS*. 326–337.