Computing Discrete Fréchet Distance

Thomas Eiter and Heikki Mannila

Christian Doppler Labor für Expertensyteme
Technische Universität Wien
Paniglgasse 16, 1040 Wien

E-mail: `eiter@vexpert.dbai.tuwien.ac.at`

April 25, 1994

CD-TR 94/64

Computing Discrete Fréchet Distance*

Thomas Eiter
Information Systems Dept.
Technical University of Vienna

Heikki Mannila
Dept. of Computer Science
University of Helsinki

Abstract

The Fréchet distance between two curves in a metric space is a measure of the similarity between the curves. We present a discrete variation of this measure. It provides good approximations of the continuous measure and can be efficiently computed using a simple algorithm. We also consider variants of discrete Fréchet distance, and find an interesting connection to measuring distance between theories.

keywords: Fréchet distance, metrics, curves, polygonal curves

1 Introduction

Given two curves in a metric space, the Fréchet distance δ_F between them can be defined intuitively as follows. A man is walking a dog on a leash: the man can move on one curve, the dog on the other; both may vary their speed, but backtracking is not allowed. What is the length of the shortest leash that is sufficient for traversing both curves? The Fréchet distance is a measure of similarity between curves that takes into account the location and ordering of the points along the curves. Therefore it is often better than the well-known Hausdorff distance. This distance function was introduced by Fréchet in 1906 [6].

A fundamental study on the computational properties of the Fréchet distance was done by Alt and Godau [1]. They give an algorithm that computes the exact Fréchet distance between two polygonal curves in time $O(pq \log^2 pq)$, where p and q are the number of segments on the polygonal curves. The algorithm is fairly involved, as it uses the parametric search technique.

In this paper we describe a discrete variation of the Fréchet distance for polygonal curves. The variation is called the *coupling distance* δ_{dF}. It is based on the idea of

*Authors' addresses: Thomas Eiter, Information Systems Department, Technical University of Vienna, Paniglgasse 16, A-1040 Vienna, Austria; Heikki Mannila, Department of Computer Science, University of Helsinki, P.O. Box 26 (Teollisuuskatu 23), SF-00014 University of Helsinki, Finland. email: `eiter@vexpert.dbai.tuwien.ac.at`, `mannila@cs.helsinki.fi`

looking at all possible couplings between the end points of the line segments of the polygonal curves.

We show that δ_{dF} provides good approximations to δ_F. Specifically, we show that δ_{dF} is an upper bound for δ_F, and that the difference between these measures is bounded by the length of the longest edge of the polygonal curves. We also show that δ_{dF} can be computed in $O(pq)$ time using a very simple algorithm.

On the basis of these results, the following way of approximately computing the Fréchet distance between two arbitrary curves suggests itself: First compute proper polygonal approximations to the curves and then compute their coupling distance, rather than computing their exact Fréchet distance.

We also briefly address variants of the discrete Fréchet distance, and find an interesting relation to a measure of distance between logical theories.

2 Discrete Fréchet distance

Following [1], we define a *curve* as a continuous mapping $f : [a, b] \to V$, where $a, b \in \Re$ and $a \leq b$ and (V, d) is a metric space.

Given two curves $f : [a, b] \to V$ and $g : [a', b'] \to V$, their *Fréchet distance* is defined as

$$\delta_F(f, g) = \inf_{\alpha,\beta} \max_{t\in[0,1]} d(f(\alpha(t)), g(\beta(t))),$$

where α (resp. β) is an arbitrary continuous nondecreasing function from $[0, 1]$ onto $[a, b]$ (resp. $[a', b']$).

In computing the Fréchet distance between arbitrary curves one typically approximates the curves by polygonal curves. A *polygonal curve* is a curve $P : [0, n] \to V$, where n is a positive integer, such that for each $i \in \{0, 1, \ldots, n-1\}$, the restriction of P to the interval $[i, i+1]$ is affine, that is $P(i + \lambda) = (1 - \lambda)P(i) + \lambda P(i + 1)$.

Let $P : [0, n] \to V$ be a polygonal curve. We denote the sequence $(P(0), P(1), \ldots, P(n))$ of endpoints of the line segments of P by $\sigma(P)$. Let P and Q be polygonal curves and $\sigma(P) = (u_1, \ldots, u_p)$ and $\sigma(Q) = (v_1, \ldots, v_q)$ the corresponding sequences. A *coupling* L between P and Q is a sequence

$$(u_{a_1}, v_{b_1}), (u_{a_2}, v_{b_2}), \ldots, (u_{a_m}, v_{b_m})$$

of distinct pairs from $\sigma(P) \times \sigma(Q)$ such that $a_1 = 1$, $b_1 = 1$, $a_m = p$, $b_m = q$, and for all $i = 1, \ldots, q$ we have $a_{i+1} = a_i$ or $a_{i+1} = a_i + 1$, and $b_{i+1} = b_i$ or $b_{i+1} = b_i$. Thus, a coupling has to respect the order of the points in P and Q. The *length* $||L||$ of the coupling L is the length of the longest link in L, that is,

$$||L|| = \max_{i=1,\ldots,m} d(u_{a_i}, v_{b_i}).$$

Given polygonal curves P and Q, their *discrete Fréchet distance* is defined to be

$$\delta_{dF}(P, Q) = \min\{||L|| \mid L \text{ is a coupling between } P \text{ and } Q\}.$$

It is immediate that $\delta_{dF}(P, P) = 0$, $\delta_{dF}(P, Q) = \delta_{dF}(Q, P)$; furthermore, one can check that $\delta_{dF}(P, Q) = 0$ implies $P = Q$ and that $\delta_{dF}(P, Q) \leq \delta_{dF}(P, R) + \delta_{dF}(R, Q)$. We thus have the following.

Proposition 1 *δ_{dF} defines a metric on the set of polygonal curves.*

The relationship of δ_{dF} to δ_F is captured by the following two lemmata, from which we immediately obtain the quality of approximation.

Lemma 2 *For all polygonal curves P and Q we have*

$$\delta_F(P, Q) \leq \delta_{dF}(P, Q).$$

Proof. A coupling with maximal edge r gives a way of walking around P and Q with leash at most r. □

Let for any polygonal curve $P = (u_1, \ldots, u_p)$ denote $D(P) = \max_{i=2,\ldots,p} d(u_{i-1}, u_i)$.

Lemma 3 *Let $P : [0, n] \to V$ and $Q : [0, m] \to V$ be polygonal curves. Then,*

$$\delta_{dF}(P, Q) \leq \delta_F(P, Q) + \max\{D(P),\, D(Q)\}.$$

Proof. Let $\alpha(t)$ (resp. $\beta(t)$) be a continuous nondecreasing function from $[0, 1]$ to $[0, n]$ (resp. $[0, m]$). Let $\sigma(P) = (u_1, \ldots, u_p)$ and $\sigma(Q) = (v_1, \ldots, v_q)$. For each point $u \in \sigma(P)$ let $t(u) \in [0, 1]$ be the smallest value such that $\alpha(t(u)) = u$, and for each point $v \in \sigma(Q)$ let $s(v) \in [0, 1]$ be the smallest value such that $\beta(s(v)) = v$.

We construct a coupling R between P and Q iteratively. First add edge (u_1, v_1) to R.

Assume then that R is already a coupling between the sequences $(u_1, \ldots, u_i)$ and $(v_1, \ldots, v_j)$. We extend R by one edge as follows. If $j = q$ or $t(u_{i+1}) < s(v_{j+1})$, then add the link (u_{i+1}, v_j). The endpoint v_j of this link is the left endpoint of the line segment in which the point $\beta(t(u_{i+1}))$ lies. Thus for the length of this link we have the inequality

$$d(u_{i+1}, v_j) \leq d(\alpha(t(u_{i+1})), \beta(t(u_{i+1}))) + D(Q). \tag{1}$$

If $i = p$ or $t(u_{i+1}) > s(v_{j+1})$, then add the link (u_i, v_{j+1}). Again, the endpoint u_i is the left endpoint of the line segment in which the point $\alpha(s(v_{j+1}))$ lies, and we have the inequality

$$d(u_i, v_{j+1}) \leq d(\alpha(s(v_{j+1})), \beta(s(v_{j+1}))) + D(Q). \tag{2}$$

Otherwise, we have $t(u_{i+1}) = s(v_{j+1})$, and we add the link (u_{i+1}, v_{j+1}), whose length is bounded by

$$d(\alpha(t(u_{i+1}), \beta(t(u_{i+1})))).$$

It is easy to see that the constructed sequence R is a coupling. The above inequalities 1 and 2 give us that

$$||R|| \leq \max_{t \in [0,1]} d(\alpha(t), \beta(t)) + \max\{D(P), D(Q)\}.$$

Since α and β were arbitrary, we obtain

$$||R|| \leq \delta_F(P,Q) + \max\{D(P),\, D(Q)\}. \quad \square$$

Theorem 4 *For any polygonal curves P and Q*

$$\delta_F(P,Q) \leq \delta_{dF}(P,Q) \leq \delta_F(P,Q) + \max\{D(P),\, D(Q)\}.$$

From this, we can regard the coupling measure δ_{dF} as a discrete version of the Fréchet-distance. More precisely, say that a polygonal curve Q is a *refinement* of the polygonal curve P, if $\sigma(P) = (u_1, \ldots, u_n)$ and for some $i = 1, \ldots, n$ we have $\sigma(Q) = (u_1, \ldots, u_i, v, u_{i+1}, \ldots, u_n)$, and the point v is on the line segment between u_i and u_{i+1}. Then,

Proposition 5 *Let $P_0, P_1, \ldots$ and $Q_0, Q_1, \ldots$ be sequences of polygonal curves such that P_{i+1} (resp. Q_{i+1}) is a refinement of P_i (resp. Q_i) for all $i \geq 0$ and $\lim_{i\to\infty} D(P_i) = \lim_{i\to\infty} D(Q_i) = 0$. Then,*

$$\lim_{i\to\infty} \delta_{dF}(P_i, Q_i) = \delta_F(P_0, Q_0).$$

Proof. By Theorem 4. □

In fact, from the results in [1] it can be seen that for any polygonal curves $P = (u_1, \ldots, u_p)$ and $Q = (v_1, \ldots, v_q)$, there always exist sequences of refinements of P and Q leading to curves P' and Q', respectively, that both contain at most $p+q$ points and satisfy $\delta_{dF}(P', Q') = \delta_F(P, Q)$.

3 Computation

An advantage of the coupling measure is its efficient computability by dynamic programming, without need of complicated data structures. The algorithm **dF** in Table 1 can be coded easily; the following lemma on its output is straightforward.

Lemma 6 **dF**$(P,Q) = \delta_{dF}(P,Q)$ *for any polygonal curves P and Q.*

Thus, we obtain the following result.

Theorem 7 *Let $P : [0,n] \to V$ and $Q : [0,m] \to V$ be polygonal curves. Denote $\sigma(P) = (u_1, \ldots, u_p)$ and $\sigma(Q) = (v_1, \ldots, v_q)$. The measure $\delta_{dF}(P,Q)$ can be computed in $O(pq)$ time.*

Proof. By Lemma 6, **dF**(P,Q) computes $\delta_{dF}(P,Q)$. It is easy to see that the runtime of **dF**(P,Q) is $O(pq)$; hence the result. □

We remark that an algorithm of Godau for deciding $\delta_F(P,Q) \leq \epsilon$? [7, 1] may be used to approximate $\delta_F(P,Q)$ on a particular machine by doing a binary search over the range r of reals that the computer normally handles. This amounts to an $O(pq \log r)$ time algorithm for fixed range of reals, which works fine for practical purposes. However, the algorithm is much more involved than the simple algorithm **dF** from above.

```
Function dF(P, Q): real;
  input:    polygonal curves P = (u_1, ..., u_p) and Q = (v_1, ..., v_q).
  return:   δ_dF(P, Q)

  ca : array [1..p, 1..q] of real;

  function c(i, j): real;
    begin
       if ca(i, j) > −1 then return ca(i, j)
       elsif i = 1 and j = 1 then ca(i, j) := d(u_1, v_1)
       elsif i > 1 and j = 1 then ca(i, j) := max{ c(i − 1, 1), d(u_i, v_1) }
       elsif i = 1 and j > 1 then ca(i, j) := max{ c(1, j − 1), d(u_1, v_j) }
       elsif i > 1 and j > 1 then ca(i, j) :=
            max{ min(c(i − 1, j), c(i − 1, j − 1), c(i, j − 1)), d(u_i, v_j) }
       else ca(i, j) = ∞
    return ca(i, j);
    end; /* function c */

  begin
    for i = 1 to p do for j = 1 to q do ca(i, j) := −1.0;
    return c(p, q);
  end.
```

Table 1: Algorithm computing the coupling measure

4 Variants of discrete Fréchet distance

A variant of discrete Fréchet distance is obtained if the length of a coupling L between P and Q is measured by the *sum* of the links in L (denote this by $||L||_s$), i.e.,

$$||L||_s = \sum_{i=1,\dots,m} d(u_{a_i}, v_{b_i}).$$

The implied measure $\delta_{sdF}(P, Q)$ gives the minimum of the total distance of an order-preserving correspondence between points of P and Q, such that each point of P corresponds to at least one point of Q and vice versa. A dog-and-leash interpretation of this measure would be the minimum total length of leashes needed for men walking their dogs, such that all points on the routes P and Q are occupied and leashes do not cross over.

Clearly, $\delta_{sdF}(P, P) = 0$, $\delta_{sdF}(P, Q) = \delta_{sdF}(Q, P)$, and that $\delta_{sdF}(P, Q) = 0$ implies $P = Q$; the triangle inequality, however, fails in general. It is easy to see that δ_{sdF} is computed by the variant of algorithm **dF** in Table 1 obtained by replacing "max" operations with additions; hence, δ_{sdF} can be computed in time $O(pq)$.

It is interesting to note a connection between measuring the distance of curves and the distance of logical theories. If one gives up on order-preservation of couplings, then the measures defined as δ_{dF} and δ_{sdF} resemble two measures for distance between logical theories. The motivation for studying such measures comes from the notion of truthlikeness in philosophy of science [11] and from artificial intelligence, where the task of quantifying the distance of theories is important for example for machine learning and theory approximation [10, 13, 12, 8, 3, 4]. Theories viewed as sets of models naturally correspond to curves viewed as sets of points. Given a metric on the models, the distance between the sets of models may be interpreted as a measure of similarity between theories [11].

Let an *unordered coupling* between P and Q be any sequence $U = (u_{a_1}, v_{b_1}), (u_{a_2}, v_{b_2}), \ldots, (u_{a_m}, v_{b_m})$ of distinct pairs from $\sigma(P) \times \sigma(Q)$ such that every point in P (resp. Q) occurs among the u_{a_i} (resp. v_{b_i}), and let $||U||$, $||U||_s$ be defined analogous as in the case of ordered couplings. Define

$$\begin{aligned}\delta^u_{dF}(P,Q) &= \min\{||U|| \mid U \text{ is an unordered coupling between } P \text{ and } Q\},\\ \delta^u_{sdF}(P,Q) &= \min\{||U||_s \mid U \text{ is an unordered coupling between } P \text{ and } Q\}.\end{aligned}$$

Both measures have been proposed for measuring theory distance. Notice that δ^u_{dF} collapses with the Hausdorff distance. The measure δ^u_{sdF} describes the minimum cost of a correspondence between the points of P and Q; it can be seen in the spirit of the unordered similarity measures in [2]. This measure, called *link measure* in the context of [5], does not define a metric, since the triangle does not hold in general.

Intuitively, the computation of δ^u_{sdF} seems to be involved, and one might suspect NP-hardness of this problem. Somewhat surprising, however, is that δ^u_{sdF} can be computed in polynomial time by using graph matching techniques (cf. [9] for a background).

Given P, $\sigma(P) = (u_1, \ldots, u_p)$ and Q, $\sigma(Q) = (v_1, \ldots, v_q)$, define a complete bipartite graph $G_0 = (A \cup B, E)$, where $A = \{a_1, \ldots, a_p\}$ and $B = \{b_1, \ldots, b_q\}$, in which each edge $e = \{a_i, b_j\}$ has weight $w(e) = d(u_i, v_j)$. Let G_1 be a zero-weight copy of G_0, and let G be the graph obtained from $G_0 \cup G_1$ by connecting each node from G to its copy by weight equal to its nearest neighbor in G. Then, the following property of G holds.

Theorem 8 *Let P and Q be polygonal curves. The cost of a minimum perfect matching in the graph G for P and Q is identical to $\delta^u_{sdF}(P,Q)$.*

(A proof can be found in [5].) Consequently, δ^u_{sdF} can be efficiently computed by applying a minimum cost perfect matching algorithm, which is feasible in polynomial time (cf. [9]).

5 Conclusion

We have presented a discrete variant of the Fréchet distance between curves in a metric space, and we described a simple and efficient algorithm for computing this measure. Besides its own interest, discrete Fréchet distance may be used for approximately computing the Fréchet distance between two arbitrary curves, as an alternative to using the exact Fréchet distance between a polygonal approximation of the curves or an approximation of this value.

Moreover, we found an interesting connection between distance of curves and logical theories. This connection suggests that distance measures for curves (sets of points) may be fruitfully applied to logical theories (sets of models).

Acknowledgements We thank Michael Godau and Hannu Toivonen for useful comments on this paper.

References

[1] H. Alt and M. Godau. Measuring the resemblance of polygonal curves. In *Proc. 8th Annu. ACM Sympos. Comput. Geom.*, pages 102–109, 1992.

[2] H. Alt, K. Mehlhorn, H. Wagener, and E. Welzl. Congruence, similarity and symmetries of geometric objects. *Discrete Comput. Geom.*, 3:237–256, 1988.

[3] M. Cadoli. Semantical and Computational Aspects of Horn Approximations. In *Proceedings IJCAI-93*, pages 39–44, Chambery, France, August 1993.

[4] M. Cadoli and M. Schaerf. Tractable Reasoning via Approximation. *Artificial Intelligence*, 1993. To appear.

[5] T. Eiter and H. Mannila. Theory Distance and Similarity: Measures and Computation. Technical Report CD-TR 94/60, CD-Laboratory for Expert Systems, TU Vienna, Austria, February 1994.

[6] M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Mathematico di Palermo*, 22:1–74, 1906.

[7] M. Godau. A natural metric for curves – computing the distance for polygonal chains and approximation algorithms. In *Proc. 8th Sympos. Theor. Aspects of Comp. STACS*, LNCS vol. 480, pages 127–136, 1991.

[8] R. Greiner and D. Schuurmans. Learning Useful Horn Approximations. In *Proceedings of the Third International Conference on Knowledge Representation and Reasoning (KR-92)*, pages 383–392, 1992.

[9] K. Mehlhorn. *Graph Algorithms and NP-Completeness*, volume 2 of *Data Structures and Algorithms*. Springer, 1984.

[10] B. K. Natarajan. *Machine Learning: A Theoretical Approach.* Morgan Kaufmann, San Mateo, CA, May 1991.

[11] I. Niiniluoto. *Truthlikeness.* D. Reidel Pub. Comp., Dordrecht, Holland, 1987.

[12] B. Selman and H. Kautz. Knowledge Compilation Using Horn Approximations. In *Proceedings AAAI-91*, pages 904–909, Anaheim, August 1991.

[13] S. Wrobel. On the Proper Definition of Minimality in Specialization and Theory Revision. In P. B. Brazdil, editor, *Machine Learning: ECML-93. European Conference on Machine Learning*, Lecture Notes in AI 667, pages 65–82. Springer-Verlag, 1993.