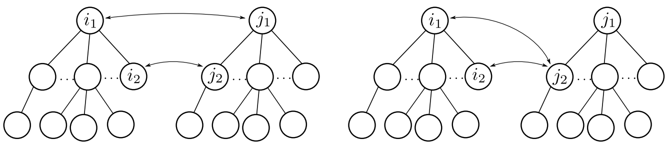
Supplementary material for "Edit Distance between Merge Trees"

Raghavendra Sridharamurthy, *Student Member, IEEE*, Talha Bin Masood, Adhitya Kamakshidasan, and Vijay Natarajan, *Member, IEEE*

Abstract—This document presents additional material supporting the paper "Edit Distance between Merge Trees". We provide examples to illustrate different properties of edit distance mappings, describe the algorithm for computing tree edit distance for constrained labeled unordered trees, and provide further details of the case study on periodicity in the 2D Bénard von Kárman vortex street dataset to show the utility of the distance measure.



(a) one-one mapping (b) violates one-one mapping

Fig. 1. Unconstrained tree edit distance mappings satisfying the one-to-one mapping property. (a) A mapping that satisfies the property. (b) A mapping that violates the property.

1 Tree edit distance mappings and illustrations

We restate the properties of tree edit distance mappings, unconstrained and constrained, and provide examples to understand their properties. For simplicity, we consider trees with equal number of nodes and that are similar to each other. Further to avoid clutter, we highlight only those mappings between the pairs of nodes that are required for the illustration of the particular property.

1.1 Unconstrained tree edit distance mappings

From Section 3.1, we know that unconstrained edit distance mappings satisfy the following properties. A triple (M_e, T_1, T_2) defines the *edit distance mapping* from T_1 to T_2, where each pair $(i_1, j_1), (i_2, j_2) \in M_e$ satisfies the following properties:

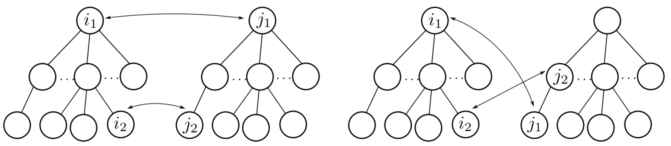
1) $i_1 = i_2$ if and only if $j_1 = j_2$ (one-to-one)
2) $t_1[i_1]$ is an ancestor of $t_1[i_2]$ if and only if $t_2[j_1]$ is an ancestor of $t_2[j_2]$ (ancestor ordering).

Figure 1 and 2 illustrate these properties using a small example. The mapping in Figure 2(b) is one-to-one but does not satisfy the ancestor preservation property, i_1 is ancestor of i_2 but j_1 is child of j_2.

1.2 Constrained tree edit distance mappings

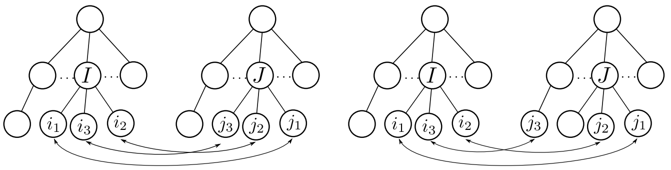
From Section 3.2 we know that constrained edit distance mappings satisfy the following properties. A triple (M_c, T_1, T_2) is called a *constrained edit distance mapping* if,

1) (M_c, T_1, T_2) is an edit distance mapping, and
2) Given three pairs $(i_1, j_1), (i_2, j_2), (i_3, j_3) \in M_c$, the *least common ancestor* $lca(t_1[i_1], t_1[i_2])$ is a proper ancestor of $t_1[i_3]$ if and only if $lca(t_2[j_1], t_2[j_2])$ is a proper ancestor of $t_2[j_3]$.



(a) ancestor preserving mapping (b) ancestor order not preserved

Fig. 2. Unconstrained tree edit distance mappings satisfying the ancestor preservation property. (a) A mapping that satisfies the property. (b) A mapping that violates the property.



(a) disjoint subtrees map to disjoint subtrees (b) disjoint subtrees do not map to disjoint subtrees

Fig. 3. Constrained tree edit distance mappings satisfying the disjoint subtree mapping property. (a) A mapping that satisfies the property. (b) A mapping that violates the property.

Figure 3 illustrates an important property required for a mapping to be constrained, namely disjoint subtrees map to disjoint subtrees. Figure 3(b) illustrates a mapping that satisfies the properties of unconstrained tree edit distance mapping but is not a constrained tree edit distance mapping. The node i_3 is a descendant (immediate descendant in this case) of the $lca(i_1, i_2) = I$ but j_3 is not a descendant of the $lca(j_1, j_2) = J$.

2 Tree edit distance algorithm for constrained labeled unordered trees

We describe the algorithm by Zhang [1] here for the sake of completeness. Algorithm 1 computes the tree edit distance. It is a dynamic programming based algorithm that follows from the properties discussed in Section 3 of the paper. Line 2 initializes the distance between two empty trees to 0. The loops spanning lines $3-6$ and $7-10$ fill the table entries corresponding to the distances between the empty tree and all trees and forests. The nested loops spanning lines $11-16$ fill the entries that correspond to distances between non-empty forests and trees. The entry $D(T_1[m], T_2[n])$ in the table with $m = |T_1|$ and $n = |T_2|$ corresponds to the final result. The algorithm computes the distance in

$$O(|T_1| \times |T_2| \times (deg(T_1) + deg(T_2)) \times log_2(deg(T_1) + deg(T_2)))$$

time in the worst case. The analysis is as described by Zhang [1]. This algorithm has a better running time than the one proposed by Xu [2].

Algorithm 1: TreeEditDistance [1]

Data: Merge trees T_1, T_2.
Result: $D(T_1[i], T_2[j])$, where $1 \leq i \leq |T_1|$, $1 \leq j \leq |T_2|$

begin
  $D(\theta, \theta) = 0$
  for $i = 1$ **to** $|T_1|$ **do**
    $D(F_1[i], \theta) = \sum_{k=1}^{n_i} D(T_1[i_k], \theta)$
    $D(T_1[i], \theta) = D(F_1[i], \theta) + \gamma(t_1[i] \longrightarrow \lambda)$
  end
  for $j = 1$ **to** $|T_2|$ **do**
    $D(\theta, F_2[j]) = \sum_{k=1}^{n_j} D(\theta, T_2[j_k])$
    $D(\theta, T_2[j]) = D(\theta, T_2[j]) + \gamma(\lambda \longrightarrow t_2[j])$
  end
  for $i = 1$ **to** $|T_1|$ **do**
    for $j = 1$ **to** $|T_2|$ **do**
      $D(F_1[i], F_2[j]) =$

$$\min \begin{cases} D(\theta, F_2[j]) + \min\limits_{1 \leq t \leq n_j} \{D(F_1[i], F_2[j_t]) - D(\theta, F_2[j_t])\}, \\ D(F_1[i], \theta) + \min\limits_{1 \leq s \leq n_i} \{D(F_1[i_s], F_2[j]) - D(F_1[i_s], \theta)\}, \\ \min\limits_{MM(i,j)} \gamma(MM(i,j)). \end{cases}$$

      $D(T_1[i], T_2[j]) =$

$$\min \begin{cases} D(\theta, T_2[j]) + \min\limits_{1 \leq t \leq n_j} \{D(T_1[i], T_2[j_t]) - D(\theta, T_2[j_t])\}, \\ D(T_1[i], \theta) + \min\limits_{1 \leq s \leq n_i} \{D(T_1[i_s], T_2[j]) - D(T_1[i_s], \theta)\}, \\ D(F_1[i], F_2[j]) + \gamma(t_1[i] \longrightarrow t_2[j]). \end{cases}$$

    end
  end
end

3 Periodicity in Time-Varying Data

Figure 4 shows few time steps of the 2D Bénard-von Kármán vortex street dataset [3] used in the periodicity case study described in Section 5.3. Figure 5 shows the full distance matrix for the case study to show that the periodic behavior is consistent across all the time steps.

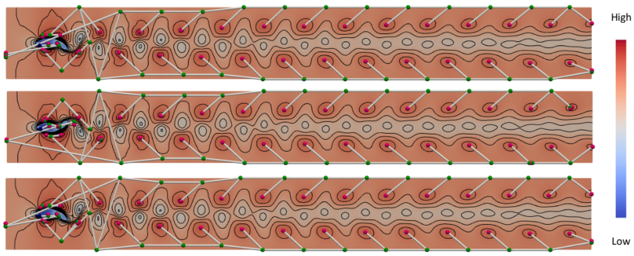


Fig. 4. Time step 0 (top), 37 (middle) and 74 (bottom) of the flow around a cylinder simulation. The split tree and the critical points are overlaid.

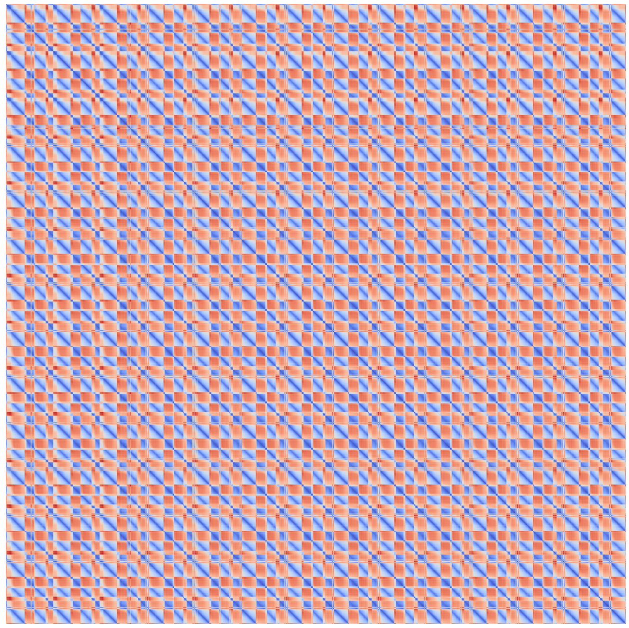


Fig. 5. The full 1000×1000 distance matrix shows a half-period of 37 in addition to the periodicity of 74-75.

References

[1] K. Zhang, "A Constrained Edit Distance Between Unordered Labeled Trees," *Algorithmica*, vol. 15, pp. 205–222, 1996.
[2] H. Xu, "An algorithm for comparing similarity between two trees," *arXiv preprint arXiv:1508.03381*, 2015.
[3] T. Weinkauf and H. Theisel, "Streak lines as tangent curves of a derived vector field," *IEEE Trans. Vis. Comp. Graphics*, vol. 16, no. 6, pp. 1225–1234, 2010.